A Work Project, presented as part of the requirements for the Award of a Master Degree in International Management from the NOVA – School of Business and Economics.

Impact and Implementation of Data Science for Social Good Projects

Jennifer Goerg

Master Student Number: 29935

A Project carried out on the Master in International Management Program, under the supervision of:

Leid Zejnilović, Ph.D.

03.01.2019

NOVA SCHOOL OF BUSINESS & ECONOMICS

Data Science For Social Good Europe

Summer Fellowship

Title:

Impact and Implementation of Data Science for Social Good Projects

Abstract:

The focus of this study is on assessing the success factors of data science projects within governmental institutions, as technology laggards, and on investigating the impact of these projects on these organizations. Using the Dynamic Capabilities as theoretical underpinning, and qualitative analysis of 6 expert interviews with key informants related to the Data Science for Social Good fellowship, this study extends the theory by analyzing the projects within scoping, prototyping and implementation. The findings contribute to the understanding of the resources and capabilities to execute data science projects in organizations that lack determinants of success, and impact of these projects.

Keywords (<4):

Dynamic Capabilities, Data Science Resources and Capabilities, Implementation Process, Organizational Impact

Index

1 Introduction

In recent years, the overwhelming daily amount of newly generated data from the web, human interactions, and other sources created many new opportunities to rethink how value is created and appropriated (Kim, Trimi, and Chung, 2014). With the emergence of the phenomenon ‘big data’ and its possible manifold applications, a newly minted interdisciplinary research field of ‘data science’ is often considered as paradigmatic change or field of power, for its access to data networks and specific resources but also the expertise to inform and educate (Williamson, 2017). The application of data analytics further profoundly affects organizational culture (Erevelles, Fukawa, and Swayne, 2016; Kitchin and McArdle, 2015; Fosso Wamba et al., 2015). Almost every organization in the world make use of data’s potential to improve firms’ outcomes and decision-making processes (Fernández et al., 2014). Although there are studies that show the possibilities of data analytics to improve decision-making, many organizations reported lacking results for their investments in data analytics (73%, Colas et al., 2014: 3). Wu, Buyya, and Ramamohanarao (2016) argue that those organizations lack essential factors which need to be considered to realize data analytics projects successfully. Colas et al. (2014) provided a list of resources and competencies, from low quality of data to inadequate analytical skills (Akter and Fasso Wamba, 2016) while Kim, Trimi, and Chung (2014) further defended that several challenges, including issues of complexity, security or technology, need to be overcome in order to obtain the successful usage of data. The surge of investment into data analytics, despite the challenges, is explained by the potential of data to transform businesses and make organizations more agile (Harvard Business Review, 2016).

Similar to the way firms are using data science to maximize profits, public organizations started using data analytics to promote social good, be it a more careful and effective expenditure of taxpayers’ money, a possibility to address issues of equity in service provision, or transparency and accountability, among others (Kim, Trimi, and Chung, 2014). During the last years, many organizations launched programs, like the Data Science for Social Good (DSSG) fellowship of the University of Chicago, to accelerate and support data science projects within governments, public and non-profit organizations (DSSG Chicago, N/Ab). Mostly perceived as not having data in its core, operating in low competing environments and lagging crucial resources and capabilities, governmental institutions require significant support during the transition process in order to serve the social good (Kim, Trimi, and Chung, 2014).

With the endeavor mentioned above, this paper aims to elaborate success factors of data science projects in governmental institutions and to assess the impact of executing these projects on the organizations. Hence, the research gap of analyzed substantial resources and capabilities and their impact in the transition process of data science projects can be identified, as the novel concept has previously not gained profound attention within the metrics of science literature. To address the gap, this paper makes two contributions: (i) The first is a collection and distribution of critical strategic resources and capabilities within project scoping, prototyping and implementation. (ii) The second is an examination of the impact on the institutions and an assessment of organizations' ability to build Dynamic Capabilities (DC) when executing a data science project.

The paper starts by introducing the DC Framework based on the Resource-Based View (RBV) and reviews the essential literature. Afterwards, necessary capabilities and resources from recent research for data science projects are listed and evaluated. An introduction of data science for governmental organizations, as well as related burdens and difficulties in this context, follow. The DSSG fellowship is considered as the empirical example of data science initiatives and serves as the methodological foundation. DSSG's partner projects are classified based on a gradual division of the project's process and used for qualitative content analysis. The application of a coding structure for the interviews enables an investigation and alteration of resources and capabilities and projects' impact on the organizations. After a presentation of the findings, a concluding discussion about distinctive reasons for a projects' success and the impact on the organization's resources and capabilities comes next. Finally, limitations are outlined, followed by a brief conclusion.

2 Theoretical Background and Frameworks

The following section is split into two main topics aiming to supply an outline on existing theories and frameworks which serve as a foundation for the assessment of the previously described approach.

2.1 Dynamic Capabilities

This chapter describes the underlying framework, relevance, and importance of Dynamic Capabilities (DC).

2.1.1 Resource-Based View as Foundation for Dynamic Capabilities

Collections of DC conceptualizations are accessible in various managerial studies. In 1997, Teece, Pisano, and Shuen published a study, which is retrospectively considered as one of the most influential researches of DC, explaining the framework as an enhancement of RBV (Barney, 1991). Differences between the underlying resources of the RBV and capabilities of a dynamic view based on Amit and Schoemaker (1993) are identified in the explanation of resources as "stocks of available factors that are owned or controlled by the firm," whereas capabilities "refer to a firm's capacity to deploy resources, usually in combination, using organizational processes, to effect a desired end" (Amit and Schoemaker, 1993: 35). Within the RBV, organization's competitiveness is based on unique resources which are rare, difficult to imitate, and valuable (Barney, 1991). Grant (1991) however classifies substantial resources into tangible, intangible, and personnel-based. Organizations can be conceptualized as an assembly of these resources, to create a competitive advantage by implementing new strategies (Ghasemaghaei, Ebrahimi, and Hassanein, 2018). Without any enlargement, the RBV breaks down when explaining competitive advantages in fast and unpredictable environments (e.g., Priem and Butler, 2001) since the duration of competitive advantage is inestimable in times of digital transformation (Eisenhardt and Martin, 2000). Therefore, the RBV was extended to a shifting competitive landscape through which a new framework evolved (Teece, Pisano, and Shuen, 1997).

2.1.2 Definitions of Dynamic Capabilities

Relentless and highly unpredictable changes in a knowledge-intensive time are normalities in which organizations must demonstrate their capability to cope with environmental and organizational shifts (Sambamurthy, Bharadwaj, and Grover, 2003). Therefore, the need for strong DC is notable as they "(…) are simple, experiential, unstable processes that rely on quickly created new knowledge and iterative execution to produce adaptive, but unpredictable outcomes" (Eisenhardt and Martin, 2000: 1106). Furthermore, DC are factors by which a manager can decide to "(…) integrate, build, and reconfigure internal and external competences to address rapidly changing environments" (Teece, Pisano, and Shuen, 1997: 516). Teece (2007) distinguishes them into three broad groups of functions, sensing (new opportunities and threats), seizing (new opportunities through investments) and transforming established strategies. Barreto (2009) further explained that organizations build DC through a learning process which cannot be bought, as their development is based on processes which are themselves determined by the organizations' assets. Those capabilities are path-depended and therefore might vary for specific projects. Success or failure of organizations, and therefore their projects, are a direct consequence of DC (Barreto, 2009).

2.1.3 Today's Need for Dynamic Capabilities

Being part of the digital transformation, by which the economy is irreversibly impacted, businesses are increasingly based upon data and have residuated digital processes in their cores. Therefore, data analytics is a key tool in times of digital transformation (Telegescu, 2018).

Digitally born companies, such as Amazon or Google known as the ‘Masters of big data,’ thoroughly exploit data’s full potential in order to realize evidence-based decision-making. By contrast, throughout various other types of organizations, data still does not have a decision-making function, even though the potential of (big) data is tremendous and widely acquainted. Whole markets shift towards their digital reformated selves, and entirely new markets are built based on emerging technologies. (McAfee and Brynjolfsson, 2012; Fosso Wamba et al., 2017) Thus, organizations need DC to develop, grow or sustain, while linking digital strategies and operational business (Fosso Wamba et al., 2017). To survive in this ever-changing environment, organizations require the ability to sustain by scooping nascent opportunities such as data science (Lu and Ramamurthy, 2011).

2.2 Data Science Projects

In 2017, Ghasemaghaei, Hassanein, and Turel developed a study to investigate crucial data analytics resources to execute data science projects. The categories ‘analytical tools’, ‘data’, ‘employee’s capabilities’ and ‘organizational task’, capture substantial factors, and are therefore employed for the identification of resources and competencies.

2.2.1 Resources and Competencies for Data Science Projects

The generic (big) data analytics capability can be defined as “the ability to acquire, store, process and analyze large amounts of (health) data in various forms, and deliver meaningful information to users (…)” (Wang and Hajli, 2017: 290). Prahalad and Krishnan (2002) reveal additional resources, such as the convergence of business operations and IT. Other studies support this view, as a strong IT capability advances decision-making, emphasizes a better reaction to changing environments (Lu and Ramamurthy, 2011) and improves digital strategies, mainly in synergy with other capabilities (Bharadwaj, 2000; Wang and Hajli, 2017). A predictive capability emphasizes the diagnosis of the future by establishing models and estimations out of large data sets to best inform the decision-making (Wessler, 2013).

Ghasemaghaei, Ebrahimi, and Hassanein (2018) mention tool sophistication as a key factor, which is espoused by Bharadwaj (2000), who further declares the quantity of data as an intangible resource leading to a stimulus of data analytics, which helps to establish insights about different stakeholders. The DSSG developed the Data Maturity Framework to provide an assessment possibility for organizations regarding readiness to start tackling social problems with data science. Data and organizational readiness, integration, as well as data governance, serve as substantial factors and capture additional important factors themselves, such as accessibility, policies, and documentation. (DSSG Chicago, N/Aa) Wang, Kung, and Byrd (2018) indicate that the data governance layer has a prevailing impact on the overall analytical architecture. Besides, Kwon, Lee, and Shin (2014) mention that the results are especially dependent on data quality. The human component of domain knowledge and analytical skills are essential resources that make resources rare, difficult to imitate and valuable (Peteraf, 1993). Domain knowledge further helps to detect key attributes for more efficient data projects (Sukumar et al., 2013). Murdoch and Detsky (2013) highlight that the sense for the economic potential of data analytics is necessary, as a lack of understanding could lead to failures of projects. In his framework, Bharadwaj (2000) suggests that together with human resources, managerial IT skills can increase the ability to coordinate multifaceted activities effectively. In addition, an alignment of IT strategies with an organization's digital and business strategies (Galliers, 2011) is a crucial mechanism for a continuously changing environment.

The mentioned papers report the resources and capabilities as critical and valuable for data science, which can help organizations to sustain and to embed dynamic decision-making approaches (Ambrosini and Bowman, 2009). Looking at the protruding RBV Framework, strategic resources alone are not sufficient for a sustainable impact (Braganza et al., 2017). Therefore, strong DC should be a high priority, as R&D also reveal further application possibilities for data science (Wernerfelt, 1984). Table 1 summarizes the mentioned resources, capabilities, or critical factors related to data science from recent studies, based on the categorization by Ghasemaghaei, Hassanein, and Turel (2017).

Table 1: Research Summary of Data Science related Resources and Capabilities

Category	Important Resources/ Capabilities	Studies
Analytical Tools/IT Infrastructure	Big Data analytics competencies	Wang, Kung, and Byrd, 2018; Telegescu, 2018
	Predictive capability	Wessler, 2013; Wang, Kung, and Byrd, 2018
	Tools sophistication	Bharadwaj, 2000; Ghasemaghaei, Ebrahimi, and Hassanein, 2018
Data	Sources	Wang and Hajli, 2017
	Governance	Wang, Kung, and Byrd, 2018; DSSG Chicago, N/Aa
	Quality	Bharadwaj, 2000; Kwon, Lee, and Shin, 2014; DSSG Chicago, N/Aa
	Bigness	Bharadwaj, 2000
	Accessibility	DSSG Chicago, N/Aa
	Aggregation layer/storage	Wang, Kung, and Byrd, 2018; DSSG Chicago, N/Aa
	Integration	DSSG Chicago, N/Aa
	Documentation	DSSG Chicago, N/Aa
	Use policy	DSSG Chicago, N/Aa
Employees Competencies/ Skills	Analytical skills	Bharadwaj, 2000; Wang and Hajli, 2017; Waller and Fawcett, 2013
	Technical IT	Bharadwaj, 2000
	Domain	Sukumar et al., 2013; Bharadwaj, 2000; Waller and Fawcett, 2013
Organizational /Managerial Factors	Sense for the economic potential of data analytics	Murdoch and Detsky, 2013
	Alignment between IT/data and business strategies	Galliers, 2011; Reich and Benbasat, 1996
	Alignment of resource commitment and strategy	Smith and Lewis, 2011
	Decision support capability	Wang, Kung, and Byrd, 2018
	Managerial IT skills	Bharadwaj, 2000
	R&D	Wernerfelt, 1984
	People resources	DSSG Chicago, N/Aa
	Leadership	DSSG Chicago, N/Aa

2.2.2 Data Science in Governmental Institutions

Within an always changing momentum of technology, it is crucial to investigate which greater purpose data science can serve and how a meaningful impact can be achieved, by solving real-life problems. Not just private but also public institutions are increasingly aware of the urge to use data to understand their citizenry, mainly to implement effective interventions, policies or improvement of projects. (Kim, Trimi, and Chung, 2014) Necessities to manage the public good (Archenaa and Mary Anita, 2015), such as solutions for water or air pollution are frequent cases for which governmental institutions rely on (big) data to increase their impact (Venkatram and Geetha, 2017). Wang and Hajli (2017) disclose several benefits to reveal the reasons for the usage of data analytics, from IT infrastructural benefits, such as the reduction of system redundancy, to organizational and managerial benefits, such as insights about emerging trends. However, governmental institutions face burdens and difficulties when adapting their resources and capabilities towards a data-driven approach in order to receive the benefits (Kim, Trimi, and Chung, 2014). The MIT Sloan Management Review published a comparison of different sectors about their status quo of resources and capabilities for digital technologies. As the second least sector in data maturity, the federal public sector performs rather poorly within all selected digital qualities, such as leadership skills. (Kane et al., 2015)

2.2.3 Burdens for Governmental Institutions

Similar to the categorization of essential factors, burdens appear related to the given dimensions. Within the **analytical tools and IT infrastructure** category, governmental institutions possess difficulties, such as open-source software issues (McAfee and Brynjolfsson, 2012). Experiences show that sharing data is considered a significant challenge, since governmental institutions, as well as partnerships, usually incorporate cross-border information flow. Tools to enable language transformation and interpretation of semantic and sentiment are often costly or are missing entirely. Similar problems occur when governmental institutions need business data sets which discern in scope and scale compared to the governmental data-sets. (Kim, Trimi, and Chung, 2014) Within the category **data**, the ‘three V’s’ characteristics from big amounts of data are often challenging. Especially governments are struggling because the volume of data needs to be in line with its regulations, the velocity causes challenges for conventional IT with the speed of the data (Ohlhorst, 2013) and the variety often requires several new technologies. (Kim, Trimi, and Chung, 2014) Public organizations tend to keep their data close, as critics might occur after providing the society with new insights. Burdens are the often missing opening processes and subsequently the necessity for new monitoring and responding mechanisms. (Janssen, Charalabidis, and Zuiderwijk, 2012) Furthermore, Kim, Trimi, and Chung (2014) mention the collection of data, coming from different channels (e.g., crowdsourcing), various sources (e.g., countries, agencies, departments) in numerous formats and with entailed costs as obstacles. Data collection and usage issues are challenges, as privacy rights often obstruct data analysis. Main technology approaches, such as open source or commercial implementations can cause difficulties, rooting from the severity of the usage to an expensive price level or specialized skill-sets of employees. (Kim, Trimi, and Chung, 2014) Related to this, difficulties within **employees' capabilities** are reported, as dealing with data science requires new skillsets accelerated by the rareness of crucial expertise, e.g., for data cleaning and organizing, as well as for the design of testing and prototyping (McAfee and Brynjolfsson, 2012). Governmental **organizational/managerial** processes also involve difficulties like the management of legacy systems (McAfee and Brynjolfsson, 2012) or funding scarcity (Kim, Trimi, and Chung, 2014). A usually profound organizational history of governmental institutions molding the culture leads to a laborious and far-reaching adaptation process of new technologies (Cianni and Steckler, 2017; Fosso Wamba et al., 2017). Moreover, extensive decision-making processes are based upon the duty to incorporate several parties (Stone, 2002). As organizations aim for agile processes and faster-paced decision making, it can be determined which resources and capabilities, such as regulations, run counter to the behaviors they are trying to instill (Cianni and Steckler, 2017; Fosso Wamba et al., 2017).

3 Research Methods

To investigate the described theoretical background and to set it into practice, the DSSG fellowship serves, together with its partners' projects, as the foundation for this analysis.

3.1 Context: The Data Science for Social Good Fellowship

The DSSG aims to train fellows to support public or non-profit organizations in areas such as health, public safety, economic development and education with the application of data science (DSSG Europe, N/Aa). Therefore, the DSSG developed an annual 12-weeks summer fellowship, which has supported 64 projects since 2013 (see Appendix A) (DSSG Chicago, N/Ab). In 2017, the University of Chicago started a cooperation with the Nova School of

Business and Economics in Lisbon to build up a second location for the fellowship (DSSG Europe, N/Aa). Besides others, machine learning or artificial intelligence are tools which are used to tackle social problems, such as high unemployment rates in Portugal, where fellows developed a dynamic-modeled system to allocate resources better and to indicate risk factors for unemployment (DSSG Europe, 2018; DSSG Europe, N/Aa).

The fellowship begins with the approach of the organizations, followed by an extensive scoping process. Until today, several hundred projects started the scoping process based on criteria such as the comprehensiveness of impact, the availability of data, as well as the organizational readiness to be biased for action (DSSG Chicago, N/Aa). If the collaboration further continues, the prototyping phase begins, which is primarily supported by the DSSG. The fellows develop a prototype and aspire to open the developed algorithms in order to sustain the codes by the reuse of others. The organizations themselves are responsible for the implementation of the projects, sometimes with the help of implementation partners and fellows. (DSSG Chicago, N/Ab) Several projects have not been implemented, and various others have not been continued during the fellowship (see Figure 1). Therefore, an assessment of factors to succeed in developing data science projects may help organizations to build the required capabilities and deliver these projects for social good.

3.2 Methodological Approach

For this study, a typical data science project is considered to traverse the phases of **(1.) scoping, (2.) prototyping** and **(3.) implementation** (Fritzler, 2015). Possible outcomes of a project are either a **success** or a **failure** at any of the phases. Projects which succeeded or failed in a later stage, succeeded by default in previous stages (see Figure 1). For this analysis, success is defined as a project which reaches the next phase, whereas failure is considered as a project which is not continued. Success in the implementation phase signifies the institutionalization of a project.

A summary list of all DSSG projects has been made, with information about the problem formulation, solution approach, organization type, and applied analysis tools. An excerpt of this list is attached in Appendix A, Table 3. However, only those partners who successfully scoped their projects are listed. Afterwards, the projects have been categorized as a success or failure within the phases (1.), (2.) and (3.), with the help of Rayid Ghani (Director of DSSG Chicago), Joseph Walsh (Data Scientist at DSSG Chicago) and Leid Zejnilović (Director of DSSG Europe). Note that although some projects are categorized as a failure, success of a project could also be understood as "the help for the social good, inspiration of people for the future, training of the partner and fellows, etc. Not all projects need to possess all the given factors in order to be successful in some kind" (R6, p.54).

The papers from Wiesche (2017) and Harrison et al. (2016) constitute a foundation for the explanation of the methodological approach. As a qualitative inquiry, this research aims to be a comprehensive and holistic investigation in the area of data science. The objectives are the DSSG fellowship and its partners, while the boundaries are the selected organizations out of all partner organizations during the last six years. The scope follows a multi-case sampling approach as the study includes insights from different potential outcomes. The exploration of unique new resources and competencies, as well as an evaluation of existing research-based variables, embedded in the theoretical DC Framework, can be described as the design of the study. Semi-structured interviews based on questionnaire guidelines (see Appendices B, C), which lasted approximately 45-60 minutes, served as the main data collection process. The interview partners were sampled out of the universe of potential outcomes (see Figure 1). Together with two expert interviews with a DSSG member and a governmental advisor, the interviews with DSSG partner organizations capture knowledge from inside and outside organizations. The categorization as either a success or failure within their specific phases, the involved knowledge of previous process phases, as well as the goal to maintain an openness during the interviews, led to individually adjusted guidelines. The interviews were recorded and transcribed.

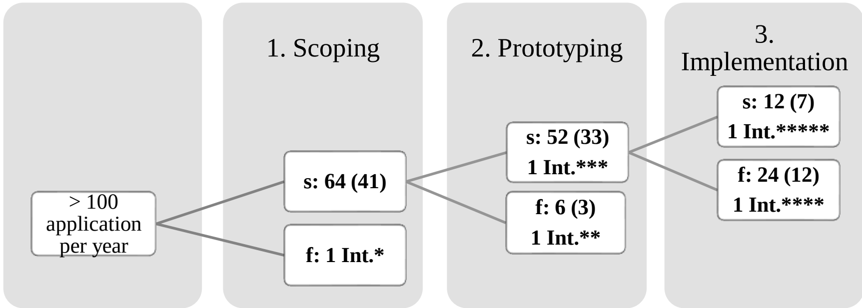

Figure 1: Outline of Data Science Project Stages with DSSG's Projects Numbers and Interviewees (derived from information in Appendix A): The outcomes 'success'(s) and 'failure'(f) are illustrated within the three phases, together with the total number of projects in each phase, the number of governmental institutions (in brackets), the number of interviewed organizations (Int.) as well as the interview partners:
* = Jamie Stainer, Principal of the Youth Training Statistic and Research Branch, Department for the Economy of the UK, (Respondent: R1)
** = Bill Thorland, Director of Evaluation and Research, Nurse-Family Partnership (Respondent: R2)
***/*/* = Tin Oreskovic, Partner for Croatian Institute of Public Health (Respondent: R3)
***** = Raed Mansour, Director, Chicago Department of Public Health (Respondent: R4)
In addition to those: Rui Lourenco, Technical Advisor, Republic of Portugal (Respondent: R5) and Joseph Walsh, Data Scientist, Center for Data Science & Public Policy & DSSG (Respondent: R6)

The analysis model of the results is based on the systematic approach of grounded theory methodology, as it is suitable for researches in the field of technological change (Wiesche, 2017). This research encompasses the explanation as a description of factors within the area of data science in governmental institutions. Therefore, this type of research can potentially be defined as a theory. After an analysis of the interviews (see Appendix D) the processes of open, axial and selective coding, combined with a constant comparison, follow (Wiesche, 2017). To capture a firms' resource and capability base, labels such as 'tool sophistication' or 'technical skills' are assigned to collect information during an open coding process. A closer investigation of collected data is done during the axial coding, while sub-themes such as 'IT infrastructure' or 'employee's capabilities,' similar to those mentioned in the theoretical background, are identified. The themes 'scoping,' 'prototyping' and 'implementation' are built during selective coding, and related codes and sub-themes are categorized accordingly. In Table 4 (see Appendix D) corresponding coding rules and descriptions are explained, followed by Table 5 (see Appendix D), which illustrates the corroborating evidence quotes from the interviewees.

Finally, a comparison of codes, sub-themes, and themes (Wiesche, 2017), reveal the similarities and differences between the necessary resources and capabilities during a data science process.

Figures 2, 3 and 4 are designed to illustrate the coding structures and to highlight the results of importance and relation between resources and capabilities to the three project phases. The figures show the frequencies of resources or capabilities mentioned by the interviewees ('**G**' (max. = 6)) and the density of correlations ('**C**'), whereas Figure 6 (see Appendix E) demonstrates the overall relations and coding structure. The results are derived from the analysis in Table 5 (see Appendix D). Resources and capabilities which are not listed in Table 1, are classified as substantial additional factors and are clustered to their respective category. To access the most important resources and capabilities in each phase and to compare the phases accordingly, the frequency of mentioned resources and capabilities is set to a cut-out value of 'G >= 2' for a summary of findings. This summary of resources and capabilities is derived in Table 2, together with the impact on organizational resources and capabilities.

4 Findings

Categorized in the phases (themes) of a data science project, the content analysis has led to the following findings and is based on the recognition that the resources and capabilities, as well as the burdens and problems mentioned in chapter 2.2, are critical factors of a project.

4.1 Scoping

Achieving success in the scoping phase is the first critical step to develop a data science project. Jamie Stainer of the Department for the Economy of the UK, whose DSSG project failed for several reasons, states that "the data opening issue is one of the main factors why we failed our project. The departments are nervous about sharing data, and the General Data Protection Regulation (GDPR) also amplifies this and lead to uncertainties" (R1, p.40). He further mentions aspects like unavailable infrastructure, data sharing burdens, data securing issues and missing human skills to cope with the burdens (R1, p.40-42). Figure 2 demonstrates the overall results for the scoping phase and reveals especially the importance of resources within the category ‘data,’ compared to the following phases (see Figure 3, Figure 4). Besides the opening process, use policies and data accessibility issues are essential factors as those are mentioned in three out of six interviews. The opening process of data, data quality, sources, and governance are solely mentioned within the scoping phase, validating the importance of the category ‘data’ (see Figure 2). Raed Mansour, Director of the Chicago Department of Public Health, states that “(…) the problem is that the data is siloed. (…) There are many databases, and they are on the federal, state and local level” (R4, p.40). ‘Opening up of data,’ ‘collaboration and partnerships,’ ‘time management,’ ‘ethical factors’ as well as ‘political factors’ are vital resources which are not mentioned in the elaborated list in chapter 2.2.

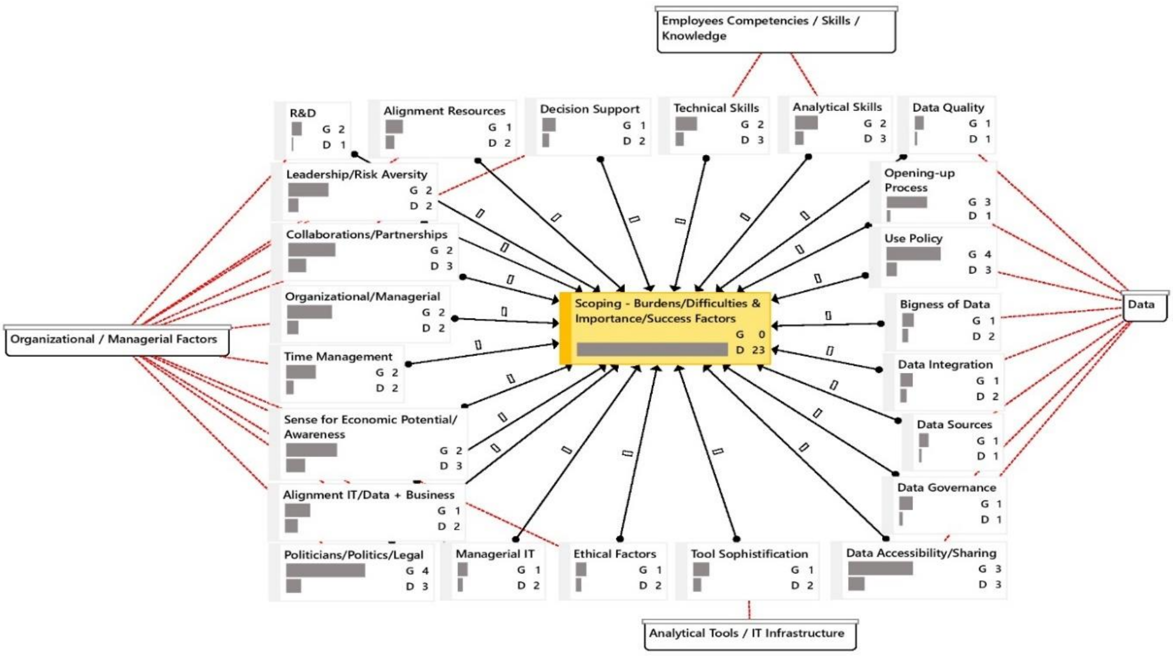

Figure 2: Scoping: Importance, Frequency & Relation of Resources & Capabilities from the Interviewees (p.40-41): The figure reveals all resources and capabilities which are reported as substantial factors within project scoping by the interviewees. All four categories are present, while especially ‘data’ and ‘organizational/managerial factors’ indicate a high influence for success.

Furthermore, the abundance of sources and administrative difficulties to effectively deal with the data is mentioned (R3, p.43). Some managerial factors are rated as especially important as, e.g., four out of six experts report politics and legal regulations (see Figure 2). IT infrastructure, as well as technical and analytical skills, are listed, even though they do not show high importance (see Figure 2). The DSSG assists the organizations during the scoping process with data revisions or general managerial resources (R3, R6, p.51).

4.2 Prototyping

The partner project ‘Nurse-Family Partnership’ failed within the prototyping phase and its expert, Bill Thorland, mentions that especially infrastructural resources, skillsets, and talent (R2, p.44) are missing to pursue new (machine learning) approaches (R2, p.43). The project of the Croatian Institute of Public Health succeeded within prototyping, and its expert, Tin Oreskovic, states that “(…) many departments and people are getting involved. The technical infrastructure is crucial, and stakeholders need to be convinced that the project is useful” (R3, p.43). The general results, illustrated in Figure 3, display that the importance of data related factors decrease, while IT infrastructure (4 out of 6 interviewees), people resources (3/6), and their skills and knowledge (e.g., general employees capabilities (3/6), domain knowledge (2/6)), increase compared to the scoping phase. The expert corroborates this, stating that “(…) there is a need for people focusing on the project during the development as they are often buried with regular report work” (R3, p.44). Funding is reported as an additional factor to the list from recent research in Table 1. The DSSG makes it to its business to strengthen the organizations with at least doubling their resources regarding data analysis during prototyping (R2, p.52). It supports the provision of resources and capabilities throughout all four categories, such as technical analysis and the establishment of predictions (R3, p.52).

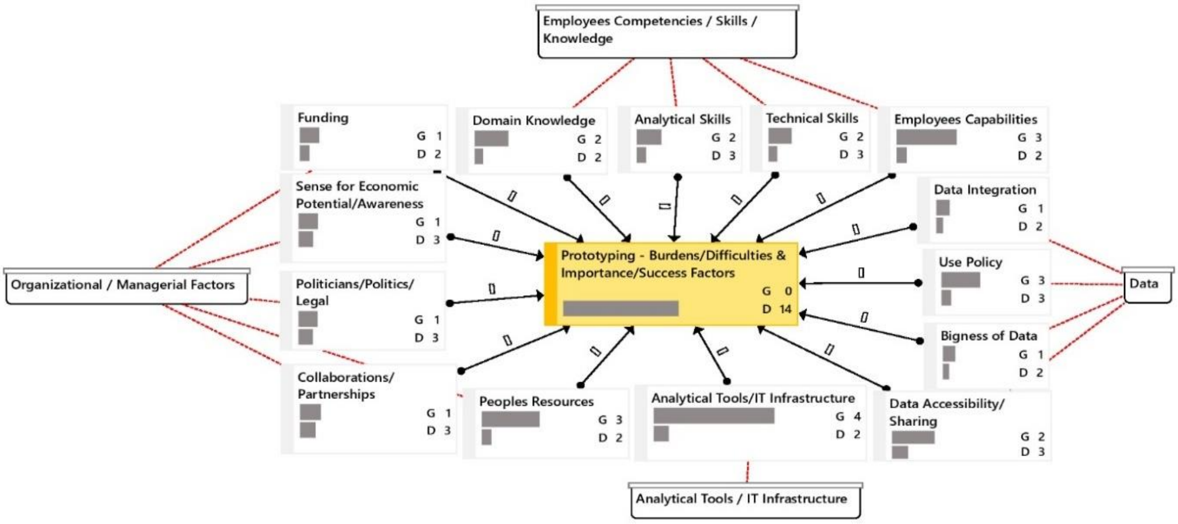

Figure 3: Prototyping: Importance, Frequency & Relation of Resources & Capabilities from the Interviewees (p.43-44): The figure reveals all resources and capabilities which are reported as substantial factors within project prototyping by the interviewees. All four categories are present, while especially 'analytical tools and IT infrastructure,' 'data' and 'employees' competencies/skills/knowledge' indicate a high influence for success.

4.3 Implementation

Within the last process step, several DSSG's partner projects succeeded and have been implemented (see Appendix A, Table 3). The expert Raed Monsur declares that "there are a lot of policies throughout all departments, which we needed to bypass when implementing the project (e.g., wage structure)" (R4, p.47). Related to this, analytical skills and additional employees' capabilities are critical as "(…) it is rather difficult to create the kind of skills which provide governments with the opportunity to participate, maintain and continue with data science projects" (R5, p.47). Figure 4 depicts the results of the implementation process and emphasizes the importance of collaborations and partnerships together with the alignment of resources (4/6 experts). Furthermore, the interviewees reveal the importance of funding as governmental institutions are mostly driven by grants, which increase the number of influencers for decision-making (R4, p.51). Within the implementation phase, additional factors like ethical concerns and communication gain importance (see Figure 4) and the factors legacy (IT) systems, additional technical requirements and communication between the departments are introduced, which enlarge the resource and capability list from recent research (see Table 1). Figure 4 highlights notably the increased influence of organizational/managerial factors and the importance of employee's competencies. Data related factors decline remarkably in concern of its importance compared to prior phases. Besides, various analytical or IT infrastructural factors are mentioned, however with a low priority (e.g., tool sophistication (1/6)). During the implementation, the DSSG supports additionally with managerial resources (R3, R6, p.52).

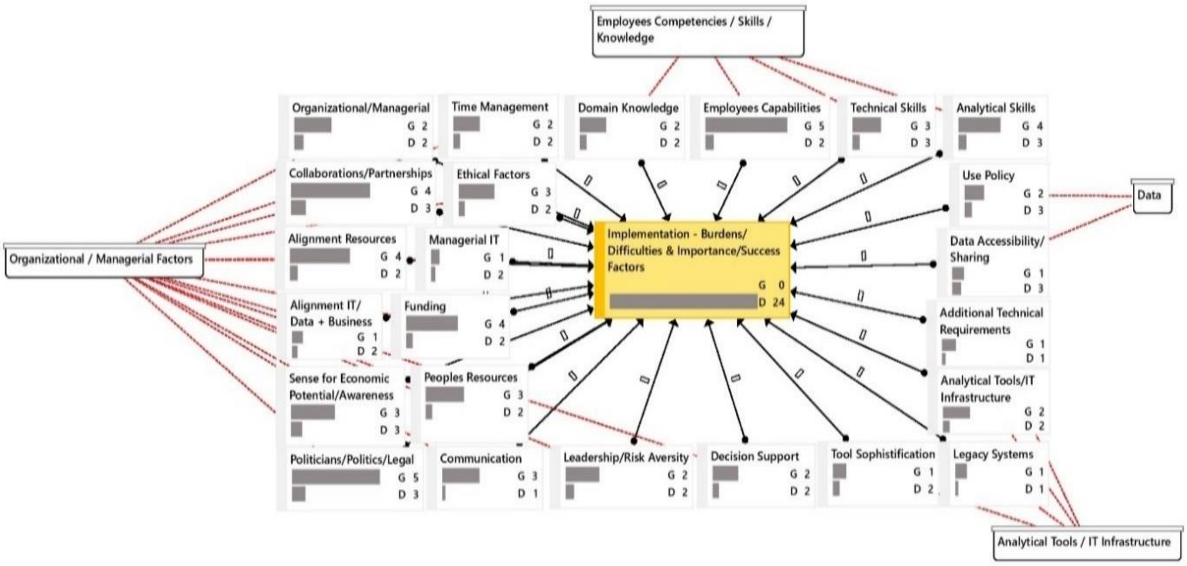

Figure 4: Implementation: Importance, Frequency & Relation of Resources & Capabilities from the Interviewees (p.46-51): The figure reveals all resources and capabilities which are reported as substantial factors within project implementation by the interviewees. All four categories are present, while especially 'organizational/managerial factors' and 'employees' competencies/skills/knowledge' indicate a high influence for success.

4.4 Impact on the Organizations

To realize data science projects, organizations firstly need resources and capabilities but are often also able to build up new, reconfigure old or learn upon resources and capabilities. Figure 5 reveals factors which are impacted and improved by the execution of a data science project according to the interviewees. Among the improvements in the communication are factors such as the alignment between organizational departments and awareness for the various possibilities of data science (e.g., R5, p.50). Also, executing a project leads to the development of processes and factors not explicitly identified in the literature review or the improvement of resources and capabilities mentioned in Table 1, such as the ability to standardize processes in order to build sustainable processes (e.g., R4, p.53), the improvement of data integration and the increase in data accessibility (e.g., R3, p.54).

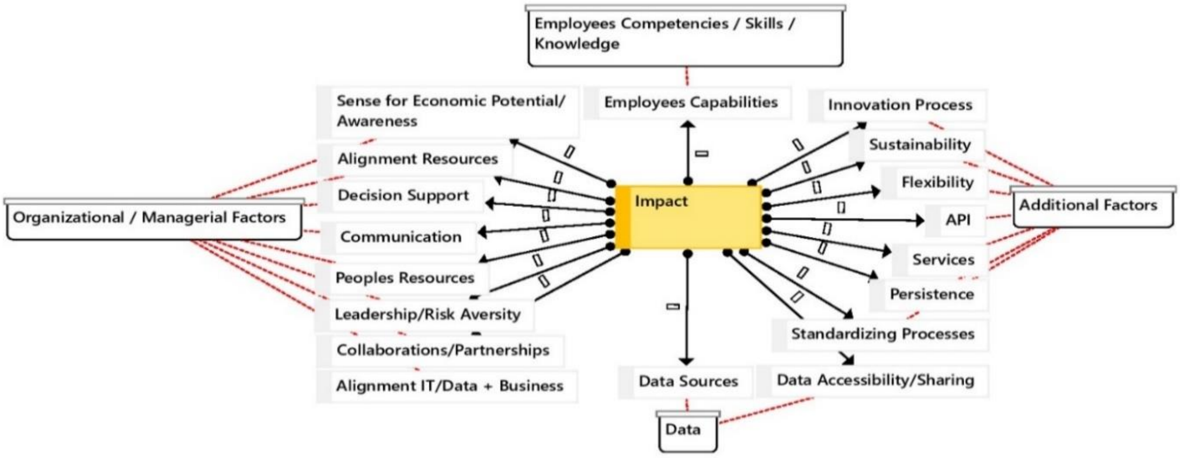

Figure 6: Impact: Importance & Relation of Resources & Capabilities to the Data Science Project Process (p.52-54): The figure reveals resources and capabilities which are reported as impacted while executing a data science project in governmental institutions by the interviewees. Especially ‘organizational/managerial factors’ are profoundly impacted by the execution of a project.

The content analysis shows that the DSSG is furthermore an influencing factor, enhancing the positive impact of data science on the organizations. As the investigated organizations are partners from the fellowship, they benefit from sustained impact with, e.g., increased sensitivity for economic potential (R1, p.52) or the need for breaking siloed databases (R4, p.53).

4.5 Summary of Findings

To identify the most pressing factors, the cutoff values for the frequency of reported resources and capabilities in project scoping, prototyping and implementation are set to ‘G>=2’. Next, each of the resources and capabilities that satisfy this condition in Figure 2, 3 and 4, are represented side-by-side in Tabe 2. Further, the table indicates those resources and capabilities which have positively been impacted by the execution of a data science project.

Table 2: Most pressing Data Science related Resources and Capabilities and Impact of the Process

Category	Important Resources/ Capabilities	Scoping	Prototyping	Implemen-tation	Impact
Analytical Tools/IT Infrastructure	Big Data analytics competencies				
	Predictive capability				
	Tools sophistication				
	General analytical tools/IT infrastructure		•	•	
Data	Sources				•
	Governance				
	Quality				
	Bigness				
	Accessibility	•	•		•
	Aggregation layer/storage				
	Integration				
	Documentation				
	Use policy	•	•	•	

Addition to Table 1	Opening process	•			
Employees Competencies/ Skills	Analytical skills	•	•	•	
	Technical IT	•	•	•	
	Domain		•	•	
	General employees' capabilities		•	•	•
Organizational /Managerial Factors	Sense for the economic potential of data analytics	•		•	•
	Alignment between IT/data and business strategies				•
	Alignment of resource commitment and strategy			•	•
	Decision support capability			•	•
	Managerial IT skills				
	R&D	•			
	People resources		•	•	•
	Leadership	•		•	•
	General organizational/managerial factors	•		•	
Addition to Table 1	Politicians/politics/legal	•		•	
	Time management	•		•	
	Collaborations/partnerships	•		•	•
	Ethical			•	
	Communication			•	•
	Funding			•	
Additional Factors and Processes influenced by the Execution of Data Science Projects	Innovation processes				•
	Sustainability of projects				•
	Flexibility in projects				•
	APIs				•
	Services of organizations				•
	Persistence of projects				•
	Standardized processes				•

5 Discussion

Based upon the theoretical evidence of crucial resources and capabilities, which have been investigated with the content analysis of interviews in four main categories, this paper validates the findings from the recent literature that various resources and capabilities are needed, but also introduces an enlarged list of essential resources and capabilities, e.g., political/legal factors or collaborations. The findings confirm that governmental institutions are facing significant challenges and are resource constraint for the execution of data science projects. The results indicate insufficient or missing resources but also difficulties with, e.g., data governance. Different resources and capabilities are crucial for each of the stages. Those that are reported as substantial from at least two interviewees 'G >=2' (see Table 2) are mainly data and managerial related factors for the scoping phase and analytical tools and IT infrastructure as well as data and employee-related factors, such as analytical skills, for the prototyping phase. The implementation phase, based on Table 2, is the most complex one, as there are many critical resources and capabilities; IT infrastructure, various employee skills and moreover numerous managerial related factors, such as alignment and decision support. Several of the identified resources and capabilities are new to the literature like ethical investigations or collaborations and partnerships, especially with the external entities (see Table 2). These findings are useful to the management of governmental institutions, as they inform where to focus attention and when.

There are notable differences among the organizations, with respect to their capabilities and resources. For example, the 'bigness' of data is a missing resource in the scoping process for one expert, while the overwhelming number of available datasets is a burden for the other interviewee. The quality of data is also reported as problematic from some experts, while others assess the quality as sufficient. (R2, p.43; R6, p.40)

Within the qualitative assessment of the impact on the organizations' resource bases, this paper unveils changes of DC amongst governmental institutions. Data accessibility and openness in the scoping phase, for example, highlight the improved organizational readiness for critical questions of the society and the willingness to interact with the environment. To build strong DC, open and accessible data would improve the governmental institutions through an increased knowledge flow (Janssen, Charalabidis, and Zuiderwijk, 2012). However, the experts mention their difficulties with opening and accessing data, and as privacy issues need to be considered for human use, which is contradicting to the crucial transparency, this might weaken DC for sensitive projects such as criminal justice (Chui et al., 2018). Either by building up capabilities such as standardizing processes, by reconfiguring resources like communication abilities or by learning technical and analytical skills (R3, p.53), DC are visible within these governmental institutions. Partnering with the DSSG augmented the ability to build DC for the organizations included in this study. For example, experts mention the influence as "(…) an impact of DSSG to learn, and build-up resources" (R3, p.52). Organizations can further learn upon resources such as sustaining processes as “(…) partners who participated twice within the fellowship, showed the ability to learn (e.g., how to build a model)” (R6, p.54).

Building DC is a balancing act. Focusing on substantial factors such as partnerships can contribute to, e.g., learning processes and reconfiguration possibilities. However, dependencies on partners can jeopardize sustainability. As partners, such as the DSSG, often support the process of prototyping, the challenges are to set up field trials and to establish mechanisms to evaluate the models independently (R6, p.54). While there is an understanding of the lacking internal expertise, strict regulations concerning the wage structure and recruiting are a big challenge. To overcome this obstacle, organizations need to reconfigure the workforce with training (R1, p.46; R4, p.47), build up educational innovation hubs (R5, p.54) or offer online training and guides (Chui et al., 2018). As the social good projects are investigated, ethical issues need to be considered. The Canadian government, for instance, is a role model for its development of a survey to assess the need for an ethics investigation when pursuing a data science project (R5, p.51). Also, political occurrences, such as the Brexit, have a formative influence on data science projects, e.g., the need in statistical and predictive systems (R1, p.45).

In general, the interviewees dissent reciprocally about the possible changes of governmental organizations towards a more agile approach through data science projects. While one expert “(…) would not say that there is currently a change in the managerial factors (…)” (R1, p.52), another argues that “(…) the outcomes can give the workforce different roles, which could be an advantage, as data is the best choice for decision-making” (R6, p.54). Governmental institutions show the ability to change their resource base and to additionally apply DC to some extent, although “(…) reconfiguring resources and being dynamic in the changing landscape is difficult, as governmental institutions are, e.g., locked in procurement regulations” (R1, p.52).

6 Limitations, Future Research, and Conclusion

This study is limited, as we know little about the conditions to develop sustainable data science projects in governmental institutions. Therefore, there might be a fourth phase of project sustaining after the implementation. Furthermore, different legal and political landscapes might be decisive for the project's outcomes, and a resulting variation of resources can influence data processing. The assumption that data science leads to better decisions might also be too simple, especially when incorporating ethical concerns. The limitation of selected domains and the focus on DSSG partners lowers the expressiveness and involves the risk of unrepresentative results. As DC usually evaluate a firm's competitive advantage, alternative frameworks for the assessment of governmental institutions could be considered, especially as the context of this paper did not allow an investigation of the new investigated assets' rarity, value, and imitability. Therefore, future research can enhance the proposed investigations in more depth and develop a guide for data science projects in governmental institutions. Notably, a future focus on the improvement of the impact on society could be derived.

As a conclusion, (i) crucial resources and capabilities differ in scoping, prototyping and implementation of data science projects in governmental institutions and (ii) the execution of data science projects support the development of DC, as new assets can be built-up, reconfigured, or acquired by learning. Partnerships, like the DSSG fellowship, are likely to have a positive impact on the organization. However, unless data science projects deliver sustained benefits and organizations can overcome the burdens, the full potential of data within governmental institutions will be limited. As the general inquiry about this topic is still in its infancy, this study is an attempt to contribute to understand its complexity and to direct future research.

7 References

Akter, Shahriar, and Samuel Fasso Wamba. 2016. “Big data analytics in E-commerce. A systematic review and agenda for future research.” *Electronic Markets*, 26(2): 173–94. https://doi.org/10.1007/s12525-016-0219-0.

Ambrosini, Véronique, and Cliff Bowman. 2009. “What are dynamic capabilities and are they a useful construct in strategic management?” *International Journal of Management Reviews*, 11(1): 29–49.

Amit, Raphael, and Paul J. H. Schoemaker. 1993. “Strategic assets and organizational rent.” *Strategic Management Journal*, 14(1): 33–46.

Archenaa, J., and E. A. Mary Anita. 2015. “A Survey of Big Data Analytics in Healthcare and Government.” *Procedia Computer Science*, 50: 408–13.

Barney, Jay. 1991. “Firm Resources and Sustained Competitive Advantage.” *Journal of Management*, 17(1): 99–120.

Barreto, Ilídio. 2009. “Dynamic Capabilities. A Review of Past Research and an Agenda for the Future.” *Journal of Management*, 36(1): 256–80.

Bharadwaj, Anandhi S. 2000. “A Resource-Based Perspective on Information Technology Capability and Firm Performance. An Empirical Investigation.” *MIS Quarterly*, 24(1): 169.

Braganza, Ashley, Laurence Brooks, Daniel Nepelski, Maged Ali, and Russ Moro. 2017. “Resource management in big data initiatives. Processes and dynamic capabilities.” *Journal of Business Research*, 70: 328–37.

Chui, Michael, Martin Harryson, James Manyika, Roger Roberts, Rita Chung, Ashley van Heteren, and Pieter Nel. 2018. “Notes from the AI Frontier applying AI for the Social Good.” McKinsey Global Institute. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx.

Cianni, Mary, and Steven Steckler. 2017. “Transforming Organizations to a Digital World.” *People & Strategy*, 40(2): 14–19.

Colas, Mathieu, Ingo Finck, Jerome Buvat, Roopa Nambiar, and Rishi R. Singh. 2014. “Cracking the data conundrum: How successful companies make big data operation.” Capgemini Consulting. https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/07/big_data_pov_03-02-15.pdf.

DSSG Chicago. N/Aa. "Data Maturity Framework" Accessed December 23. https://dsapp.uchicago.edu/home/resources/datamaturity/.

DSSG Chicago. N/Ab. "Home" Accessed December 16. https://dssg.uchicago.edu/.

DSSG Chicago. N/Ac. "Projects | Data Science for Social Good" Accessed December 28. https://dssg.uchicago.edu/projects/.

DSSG Europe. N/Aa. "DSSG Europe | Data Science for Social Good" Accessed December 24. https://dssg.uchicago.edu/europe/.

DSSG Europe. N/Ab. "Projects | Data Science for Social Good" Accessed December 28. https://dssg.uchicago.edu/europe/europe-projects/.

DSSG Europe. 2018. "Predicting long-term unemployment in Portugal" Accessed December 31. https://dssg.uchicago.edu/project/predicting-long-term-unemployment-in-continental-portugal/.

Eisenhardt, Kathleen M., and Jeffrey A. Martin. 2000. “Dynamic capabilities: what are they?” *Strategic Management Journal*, 21: 1105–21.

Erevelles, Sunil, Nobuyuki Fukawa, and Linda Swayne. 2016. “Big Data consumer analytics and the transformation of marketing.” *Journal of Business Research*, 69(2): 897–904.

Fernández, Alberto, Sara del Río, Victoria López, Abdullah Bawakid, María J. del Jesus, José M. Benítez, and Francisco Herrera. 2014. “Big Data with Cloud Computing. An insight on the computing environment, MapReduce, and programming frameworks.” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(5): 380–409.

Fosso Wamba, Samuel, Shahriar Akter, Andrew Edwards, Geoffrey Chopin, and Denis Gnanzou. 2015. “How 'big data' can make big impact: findings from a systematic review and a longitudinal case study.” *Faculty of Business - Papers*, 165: 234–46.

Fosso Wamba, Samuel, Angappa Gunasekaran, Shahriar Akter, Steven J.-f. Ren, Rameshwar Dubey, and Stephen J. Childe. 2017. “Big data analytics and firm performance. Effects of dynamic capabilities.” *Journal of Business Research*, 70: 356–65.

Fritzler, Alan. 2015. "An Ethical Checklist for Data Science | Data Science for Social Good" Accessed December 29. https://dssg.uchicago.edu/2015/09/18/an-ethical-checklist-for-data-science/.

Galliers, Robert D. 2011. *Further Developments in Information Systems Strategizing: Unpacking the Concept*. New York: Oxford University Press.

Ghasemaghaei, Maryam, Sepideh Ebrahimi, and Khaled Hassanein. 2018. “Data analytics competency for improving firm decision making performance.” *The Journal of Strategic Information Systems*, 27(1): 101–13.

Ghasemaghaei, Maryam, Khaled Hassanein, and Ofir Turel. 2017. “Increasing firm agility through the use of data analytics. The role of fit.” *Decision Support Systems*, 101: 95–105.

Grant, Robert M. 1991. “The Resource-Based Theory of Competitive Advantage. Implications for Strategy Formulation.” *California Management Review*, 33(3): 114–35.

Harrison, Helena, Melanie Birks, Richard Franklin, and Jane Mills. 2016. “Case Study Research: Foundations and Methodological Orientations.” *Qualitative Social Research*, 18(1). http://www.qualitative-research.net/index.php/fqs/article/view/2655/4079.

Harvard Business Review. 2016. “Funding IT Transformation: How Companies Can Get Past Legacy Lock-in and Stop Playing Catch-Up.” Harvard Business School Publishing. https://hbr.org/resources/pdfs/comm/hp/19804_HBR_Report_HPE.pdf.

Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. “Benefits, Adoption Barriers and Myths of Open Data and Open Government.” *Information Systems Management*, 29(4): 258–68.

Kane, Gerald C., Dough Palmer, Anh N. Phillips, David Kiron, and Natasha Buckley. 2015. “Strategy, Not Technology, Drives Digital Transformation.” MIT Sloan Management Review and Deloitte University Press. https://sloanreview.mit.edu/projects/strategy-drives-digital-transformation/.

Kim, Gang-Hoon, Silvana Trimi, and Ji-Hyong Chung. 2014. “Big-data applications in the government sector.” *Communications of the ACM*, 57(3): 78–85.

Kitchin, Rob, and Gavin McArdle. 2015. “The Diverse Nature of Big Data.” *Big Data & Society*, 3: 1–10. http://dx.doi.org/10.2139/ssrn.2662462.

Kwon, Ohbyung, Namyeon Lee, and Bongsik Shin. 2014. “Data quality management, data usage experience and acquisition intention of big data analytics.” *International Journal of Information Management*, 34(3): 387–94.

Lu, Ying, and Vazhappady K. Ramamurthy. 2011. “Understanding the Link Between Information Technology Capability and Organizational Agility: An Empirical Examination.” *MIS Quarterly*, 35(4): 931–54.

McAfee, Andrew, and Erik Brynjolfsson. 2012. “Big Data: The Management Revolution. Exploiting vast new flows of information can radically improve your company’s performance. But first you’ll have to change your decision-making culture.” 10.

Murdoch, Travis B., and Allan S. Detsky. 2013. “The inevitable application of big data to health care.” *JAMA*, 309(13): 1351–52.

Ohlhorst, Frank J. 2013. *Wiley and SAS Business Series*, *Big Data Analytics. Turning Big Data into Big Money*. New York: John Wiley & Sons.

Peteraf, Margaret A. 1993. “The cornerstones of competitive advantage. A resource-based view.” *Strategic Management Journal*, 14(3): 179–91.

Prahalad, C. K., and M. S. Krishnan. 2002. “The Dynamic Synchronization of Strategy and Information Technology.” *MIT Sloan management review*, 43(4): 24–33.

Priem, Richard L., and John E. Butler. 2001. “Is the Resource-Based "View" a Useful Perspective for Strategic Management Research?” *Academy of Management Review*, 26(1): 22–40.

Reich, Blaize H., and Izak Benbasat. 1996. “Measuring the Linkage between Business and Information Technology Objectives.” *MIS Quarterly*, 20(1): 55.

Sambamurthy, V., Anandhi S. Bharadwaj, and Varun Grover. 2003. “Shaping Agility through Digital Options. Reconceptualizing the Role of Information Technology in Contemporary Firms.” *MIS Quarterly*, 27(2): 237.

Smith, Wendy K., and Marianne W. Lewis. 2011. "Toward a Theory of Paradox. A Dynamic equilibrium Model of Organizing." *Academy of Management Review*, 36(2): 381–403.

Stone, Deborah. 2002. *Policy Paradox. The art of political decision making*. New York: Norton.

Sukumar, Sreenivas R., Regina K. Ferrell, Bonnie Carroll, and Edrick G. Coppock. 2013. "'Big Data' collaboration. Exploring, recording and sharing enterprise knowledge." *Information Services & Use*, 33(3-4): 257–70.

Teece, David J. 2007. "Explicating dynamic capabilities. The nature and microfoundations of (sustainable) enterprise performance." *Strategic Management Journal*, 28(13): 1319–50.

Teece, David J., Gary Pisano, and Amy Shuen. 1997. "Dynamic Capabilities and Strategic Management." *Strategic Management Journal*, 18(7): 509–33.

Telegescu, Tiberiu. 2018. "IT in the workspace – The need for digital transformation." *Proceedings of the International Conference on Business Excellence*, 12(1): 952–65.

Venkatram, Kari, and Mary A. Geetha. 2017. "Review on Big Data & Analytics – Concepts, Philosophy, Process and Applications." *Cybernetics and Information Technologies*, 17(2): 3–27.

Waller, Matthew A., and Stanley E. Fawcett. 2013. "Data Science, Predictive Analytics, and Big Data. A Revolution That Will Transform Supply Chain Design and Management." *Journal of Business Logistics*, 34(2): 77–84.

Wang, Yichuan, and Nick Hajli. 2017. "Exploring the path to big data analytics success in healthcare." *Journal of Business Research*, 70: 287–99.

Wang, Yichuan, LeeAnn Kung, and Terry A. Byrd. 2018. "Big data analytics. Understanding its capabilities and potential benefits for healthcare organizations." *Technological Forecasting and Social Change*, 126: 3–13.

Wernerfelt, Birger. 1984. "A resource-based view of the firm." *Strategic Management Journal*, 5(2): 171–80.

Wessler, Michael. 2013. *Big Data Analytics for Dummies*. Hoboken, NJ: John Wiley & Sons.

Wiesche, Manuel. 2017. "Grounded Theory Methodology in Information Systems Research." *MIS Quarterly*, 41(3): 685–701.

Williamson, Ben. 2017. "Who owns educational theory? Big data, algorithms and the expert power of education data science." *E-Learning and Digital Media*, 14(3): 105–22. http://fesrvsd.fe.unl.pt:2502/doi/10.1177/2042753017731238.

Wu, Caesar, Rajkumar Buyya, and Kotagiri Ramamohanarao. 2016. *Big Data Analytics = Machine Learning + Cloud Computing*. N/A: N/A.

8 Appendices

A. Excel Sheet – Summary of DSSG Projects (Excerpt)

Table 3: Summary of DSSG Projects and their Outcomes

				Process (success/failure)		
Project Name & Partner Institution & Type of Partner	**Year**	**Problem / Goal**	**Analysis/ Tools**	**Scoping**	**Prototyping**	**Implement-ation**
Simulating better bus service Chicago Transit Authority (CTA) (Non-profit)	**2013**	Help CTA to predict better the impact of a service change on a route – and all connecting routes – before deploying a single vehicle.	cutting-edge simulation	**success**	**No information available**	**No information available**
Predicting Divvy bike share stations occupancy rate Chicago Department of Transportation (Public/ Governmental Institution)	**2013**	Predicting how full or empty bike station will be in an hour or two	prediction model	**success**	**failure**	
The Giving Graph The Case Foundation (Non-profit)	**2013**	How to connect people to the causes they are passionate about	Build of a proof of concept; using social graphs	**success**	**success**	**success**
Prevent Cardiac Arrests (Code Blue) NorthShore University HealthSystem **(**Public/ Governmental Institution)	**2013**	Childhood obesity and cardiac arrest (Code blue)		**success**	**success**	**failure**
Measure the impact of early childhood health programs Nurse-Family Partnership **(**Non-profit and Public Organization)	**2013**	Reducing poverty, instability, and despair for the young woman who got pregnant before they were ready		**success**	**failure**	
Predicting building energy savings Berkeley Lab, Agentis Energy **(**Public/ Governmental Institution)	**2013**	Increase Energy Savings	using Berkeley Lab's building fingerprint tool for the prediction model	**success**	**No information available**	**No information available**
Getting kids into college Mesa Public Schools **(**Public/ Governmental Institution)	**2013**	Identifying college-ready candidates		**success**	**No information available**	**No information available**
Returning Vacant Land To Productive Use Cook County Land Bank **(**Public/ Governmental Institution)	**2013**	Returning a Vacant Land to Productive Use	Build a web interface for public use and use by land bank employees	**success**	**No information available**	**No information available**
Predictive Analytics for Smarter City Services City of Chicago, Chapin Hall (Public/ Governmental Institution)	**2013**	Upgrade System that is used now	Prediction model	**success**	**No information available**	**No information available**

Smarter crowdsourcing for crisis maps Ushahidi (Non-profit)	**2013**	Creating a report of people`s needs during a crisis	using machine learning	**success**	**No information available**	**No information available**
Measuring disaster damage with tweets Qatar Computing Research Institute (Public/ Governmental Institution)	**2013**	How to measure disaster damage to social media		**success**	**No information available**	**No information available**
Increasing the Efficiency of Intervention Strategies for the Homeless Chicago Alliance to End Homelessness	**2014**	Decrease number of homeless people		**success**	**failure**	
Collusion in International Development Projects World Bank Group (Non-profit)	**2014**	Decrease collusion and corruption risk	The model that predicts potential collusion cases	**success**	**success**	**failure**
Prevent Lead Poisoning in Children Chicago Department of Public Health (Public/ Governmental Institution)	**2014**	Preventing Lead Poisoning in Children	Building statistical models that predict exposure	**success**	**success**	**success**
Reducing Maternal Mortality Rates in Mexico Mexico's Office of the President (Public/ Governmental Institution)	**2014**	Maternal Mortality Rates	develop individual-level models of risk using all available data	**success**	**success**	**success**
Targeting of Uninsured Citizenz for the Health Insurance Enrollment Get Covered Illinois	**2014**		Construction of models for communication and messaging to reach the various subpopulations of the uninsured	**success**	**success**	**failure**
Improving Social Services Interactions Health Leads (Non-profit)	**2014**	Improving Social Services Interactions		**success**	**success**	**failure**
Predicting Success in Mother-Child Interventions Nurse-Family Partnership (Non-profit and Public Organization)	**2014**	Identify mothers who should benefit from the NFP's programs and nurse visits		**success**	**success**	**failure**
Government Spending Bills to Understand Pork Spending Sunlight Foundation, Chicago Harris (Non-profit)	**2014**	How to filter unusable data from government bills	In-depth analysis	**success**	**success**	**failure**
Sensor Data to Evaluate Environmental Initiatives Conservation International (Non-profit)	**2014**	Get most information from the sensor network	Developed algorithms for inter- and extrapolating camera trap data (generate micro-climate information)	**success**	**failure**	
Building Tools to Analyze Smart Meter Data	**2014**	Create a smarter and more efficient electric grid and create energy management	residential energy management tools	**success**	**success**	**success**

Pecan Street Inc., WikiEnergy		platforms and products that appeal to consumers				
Precicting Student Enrollments for the Allocation of Budget Chicago Public Schools (Public/ Governmental Institution)	**2014**	Better distribution of money	develop a predictive model	**success**	**failure**	
College Readiness for High School Students Montgomery County Public School District (Public/ Governmental Institution)	**2014**			**success**	**success**	**success**
Identifying Skills Gaps to Reduce Unemployment Skills for Chicagoland's Future, CareerBuilder	**2014**	Reducing unemployment		**success**	**success**	**failure**
Urban Investments for Future Economic Outcomes City of Memphis (Public/ Governmental Institution)	**2014**	help the City of Memphis develop a new system for informing policy and investment decisions		**success**	**success**	**No information available**
Identifying Opportunities for Food Bank Donation Feeding America (Non-profit)	**2015**			**success**	**success**	**failure**
Developing of Early Intervention System for Adverse Police Interactions	**2015**			**success**	**success**	**success**
Identify Fraud and Collusion in Int. Development Projects World Bank Group	**2015**	Identifying Fraud & Collusion in International Development Projects		**success**	**success**	**success**
Long-Term Financial Soundness for Home Abandonment in Mexico Infonavit (Public/ Governmental Institution)	**2015**	Enhance the quality of life and equity value of Mexicans		**success**	**success**	**failure**
Predict and Reduce Adverse Birth Outcomes	**2015**	Predicting and Reducing Adverse Birth Outcomes	prediction models	**success**	**success**	**failure**
Lobbyists Through State Legislatures Sunlight Foundation (Non-profit)	**2015**	Origin of the bills; tracing the ideas		**success**	**success**	**success**
Enforcement of Pollution and Waste Violations U.S. Environmental Protection Agency (Public/ Governmental Institution)	**2015**	To protect and improve the natural resources and environment	prediction models	**success**	**success**	**failure**
Digital Engagement for Environmental Causes with the Help of Data Australian Conservation Foundation	**2015**	Analyze data for the deeper understanding of community and for creating more effective communications	engagement models that predict which individuals are likely to take particular actions	**success**	**success**	**failure**

(Non-profit)			and the best way to communicate with those people. -> usage for experiments			
Predicting the Persistence of College from High School Students (Public/ Governmental Institution)	**2015**	Creating better methods		**success**	**success**	**failure**
Analyzing High School Students risking to not Graduate on Time Arlington , Vancouver, and Wake County Public Schools, **(**Public/ Governmental Institution)	**2015**	Identifying likelihood of High School Students not Graduating on Time	early warning systems	**success**	**success**	**failure**
Improving the Matching of the Labor Market with the use of High-Frequency Resume and Jobs Data U.S. Department of Labor **(**Public/ Governmental Institution)	**2015**	Reducing unemployment by filling skills gaps		**success**	**success**	**failure**
Proactive Blight Reduction and Neighborhood Revitalization City of Cincinnati **(**Public/ Governmental Institution)	**2015**	Prevent houses and buildings from collapsing		**success**	**success**	**success**
Identifying Frequent Users of Public Systems Johnson County, KS **(**Public/ Governmental Institution)	**2016**	Data siloed within agencies slow coordinated effort across entities to help vulnerable individuals.	prediction model	**success**	**success**	**success**
Distribution of Social Services in Mexico SEDESOL Mexico **(**Public/ Governmental Institution)	**2016**	Develop the system to help the agency improve the living conditions of poor populations in Mexico.		**success**	**success**	**failure**
Juvenile Interactions with the Criminal Justice System City of Milwaukee **(**Public/ Governmental Institution)	**2016**	Integrating data from across several different departments and systems into a unified platform called DataShare	prediction model	**success**	**success**	**failure**
Improving the quality and delivery system of Emergency Medical Services City of Cincinnati **(**Public/ Governmental Institution)	**2016**	Increase EMS efficiency, effectiveness, and long-term sustainability.		**success**	**success**	**failure**
Enforcement of Pollution and Waste Violations New York State Department of Environmental Conservation **(**Public/ Governmental Institution)	**2016**	To protect and improve the natural resources and environment in the state New York	predictive models that identify facilities with a high likelihood of violating environmental regulations	**success**	**success**	**failure**
Building up a Police Early Intervention System Charlotte-Mecklenburg Police Department	**2016**	Build a better “early intervention system” (EIS).		**success**	**success**	**failure**

(Public/ Governmental Institution)						
Intervention System for Adverse Police Interactions Metro Nashville Police Department **(**Public/ Governmental Institution)	**2016**	Expanding Our Early Intervention System		**success**	**success**	**success**
Early Warning System for Water Infrastructure Problems City of Syracuse **(**Public/ Governmental Institution)	**2016**	Build an early warning system for water infrastructure problem	prediction model	**success**	**success**	**success**
Waste Collection in Kenya from Portable Sanitation Sanergy	**2016**	Optimizing the collection of waste	prediction algorithm	**success**	**failure**	
Improving Government Response to Citizen Requests Office of the President of Mexico **(**Public/ Governmental Institution)	**2016**	Improving the responses given by the government to Citizen (online)	algorithms to classify petitions based on their content and importance, route them to the correct government agency, and partially automate the responses to common requests + machine learning	**success**	**success**	**No information available**
Predicting Students Struggling Academically by Third Grade Tulsa Public Schools **(**Public/ Governmental Institution)	**2016**	Identification of children at risk of struggling academically	develop an “early warning system.”	**success**	**success**	**failure**
Influencing Students at Risk from dropping out in High School Muskingum Valley Educational Service Center **(**Public/ Governmental Institution)	**2016**	Identifying and Influencing Students		**success**	**success**	**No information available**
Identifying rooftop usage in Rotterdam* The Municipality of Rotterdam **(**Public/ Governmental Institution)	**2017**	Identifying opportunities to use Rotterdam’s rooftops to address challenges like water storage, green spaces, and energy generation		**success**	**success**	**No information available**
Improving incident response in the Netherlands* Rijkswaterstaat **(**Public/ Governmental Institution)	**2017**	Develop a policy to optimize stationary and patrolling locations of inspectors and optimize safety and traffic flow		**success**	**success**	**No information available**
Predicting risk of long-term unemployment* Câmara Municipal de Cascais, Instituto do Emprego e Formação Profissional (Public/ Governmental Institution)	**2017**	Develop a recommendation system to increase the likelihood of people switching from unemployed to employed status		**success**	**success**	**failure**

Sustainable Tourism in Tuscany* Toscana Promozione Turistica **(**Public/ Governmental Institution)	**2017**	Measure tourism above traditional surveys and statistics to explore and design solutions for sustainable tourism in the city		**success**	**success**	**No information available**
Matchmaking between patients and doctors* José de Mello Saúde	**2017**	Increase the likelihood of long-lasting relationships, preventive medicine and quality of follow-ups		**success**	**success**	**failure**
Fishing Risk Framework developed with data of Satellites and Oceans*	**2017**	End Illegal and not reported fishing (IUU).	machine learning	**success**	**success**	**failure**
Alliance Chicago AllianceChicago	**2018**	Identify people with the risk of diabetes and to improve screening guidelines for testing people. Prevention effort to promote earlier and more efficient screening for type 2 diabetes		**success**	**success**	**In progress**
School Dropout of Social Programs in El Salvador Ministerio de Educación de El Salvador (MINED) **(**Public/ Governmental Institution)	**2018**	Reduce early school drop-out rates in El Salvador		**success**	**success**	**In progress**
Improving Workplace Safety through Proactive Inspections Labour Agency, Chile	**2018**	Better identify companies that have a high probability of infractions		**success**	**success**	**In progress**
Improving Traffic Safety Through Video Analysis Jakarta Smart City & UN Global Pulse, Jakarta **(**Public/ Governmental Institution)	**2018**	Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. Improve Jakarta's Traffic situation		**failure**	**success**	**In progress**
Recidivism for People with Health Needs Johnson County, Kansas **(**Public/ Governmental Institution)	**2018**	Identify top 200 people who are likely to get another booking within the jail and try to get them into medical help instead of jail		**success**	**success**	**In progress**
Optimizing Tourism in Tuscany* Toscana Promozione Turistica	**2018**	Measure tourism above traditional surveys and statistics to explore and design solutions for sustainable tourism in the city		**success**	**success**	**In progress**
Road accidents on highways (Netherlands)* Rijkswaterstaat **(**Public/ Governmental Institution)	**2018**	Develop a policy to optimize stationary and patrolling locations of inspectors and optimize safety and traffic flow	predictive model	**success**	**success**	**In progress**
Predicting long-term unemployment in Portugal* IEFP (Instituto de Emprego e Formação Profissional) **(**Public/ Governmental Institution)	**2018**	Develop a recommendation system to increase the likelihood of people switching from unemployed to employed status		**success**	**success**	**In progress**
School-Medicine Service for Efficient Promotion of MMR Vaccination* Croatian Institute of Public Health (CIPH) (Key	**2018**	Reduce the number of patients; increase the vaccination rate for measles, mumps, and rubella; reaching the target		**success**	**success**	**In progress**

Partner), Institute of Public health of Split-Dalmatia County, Andrija Stamper School of Public Health, Croatian Society for School and University Medicine (Public/ Governmental Institution)		vaccination rate of >95% amongst children.				

*Note. In*formation for the projects of the Data Science for Social Good fellowship Chicago from DSSG Chicago (N/Ac) and information for the projects of the Data Science for Social Good fellowship Europe projects (all projects marked '*') from DSSG Europe (N/Ab).

B. Questionnaire Guideline for DSSG Manager and Fellows

Semi-structured questionnaire – example questions:

Personal:

- What is your role within the DSSG fellowship?
- For how long are you a part of the DSSG?

General:

- How do you define 'success' of a data science project?
- How do you define 'failure' of a data science project?

Cooperation with the organizations:

- How could you evaluate the collaboration of the DSSG with the organizations before, during and after the fellowship?
- What kind of support was provided for the organizations? How do you evaluate the bias/influence of the DSSG for the projects/organizations in the following phases?
  A) Scoping B) Prototyping C) Implementation

Project / Data Science:

- How do you rate the importance of data science in governmental institutions? Which added value do you see in the area of data science for organizations?
- How do you rate the organizational readiness for digital transformation or change through data science?
- How do you rate the organizational readiness of governmental institutions for digital transformation or change through data science?
- Are governmental institutions usually lagging behind businesses within digital transformation and the adoption of data science?
- Which burden/difficulties do you see for a data science projects in the following phases?
  A) Scoping B) Prototyping C) Implementation
- Which resources/capabilities are necessary (most important) for the organizations in the following phases?
  A) Scoping B) Prototyping C) Implementation
- Which capabilities are built/improved within organizations while establishing data science projects during the following phases?
  A) Scoping B) Prototyping C) Implementation
- Does the data science project have an enabler or performer function for decision-making?
- What functions do the outcomes of data science projects have in a governmental organization? Do data science projects change the organizational structure? Do the outcomes substitute people or give them different roles?
- How would you rate ethical concerns/objections while executing a data science project and therewith the decision power of data science?
- What changes/adjustments within organizations need to be done in order to integrate data science projects or the outcomes of projects?

C. Questionnaire Guideline for DSSG Partner Institutions

<u>Semi-structured questionnaire – example questions:</u>

Note. The Interviewees have been asked specific questions to the phase they were categorized (see Figure 1). Furthermore, the Interviewee got an introduction in the topic and all necessary information about the categorization system of resources and capabilities. Therefore, this guideline was adjusted for every interviewee.

Personal:

- What is your role/job title in your organization?
- What is/was your role within the DSSG process?
  A) Scoping B) Prototyping C) Implementation

Organizational:

- How heavily does your organization invest in/work with data science?
- Does your organization have internal scoped, prototyped or implemented data science projects?

Cooperation with DSSG:

- How do you evaluate the collaboration with the DSSG before, during and after the fellowship?
- What kind of support was provided by the DSSG? How do you evaluate the bias/influence of the DSSG in the following phases?
  A) Scoping B) Prototyping C) Implementation
- Do you perceive your project as a success or failure within the project's phases?
  A) Scoping B) Prototyping C) Implementation
  Do you perceive your project as a success or failure in relation to the impact on your organization?

Project / Data Science:

- How do you rate the importance of data science in governmental institutions? Which added value do you see in the area of data science for your organization?
- How do you rate your organizational readiness for digital transformation or change through data science?
- Would you rate your organization as lagging behind other businesses within digital transformation and the adoption of data science?
- Which burden/difficulties do you see for your data science project in the following phases?
  A) Scoping B) Prototyping C) Implementation
- Which resources/capabilities are necessary for your organization in the following phases?
  A) Scoping B) Prototyping C) Implementation
- Which capabilities are built/improved within your organization while establishing data science projects during the following phases?
  B) Scoping B) Prototyping C) Implementation
- Does the project's outcome have an enabler or performer function for decision-making?

- What functions do the outcomes of data science projects have in your organization? Do data science projects change the organizational structure? Do the outcomes substitute people or give them different roles?
- How would you rate ethical concerns/objections while executing a data science project and therewith the decision power of data science?
- What changes/adjustments within your organizations need to be done in order to integrate data science projects or the outcomes of projects?

D. Qualitative Content Analysis of Interviews

Table 4: Categorization System for Content Analysis

	Categorization and Theme Description	Definition / Rule	Special Coding Rule
1	**Scoping**	All resources/capabilities which are either especially important / success factors for the scoping or in which governments are lagging/ have difficulties (also considered as indicators for critical factors)	Especially attention to the interviewees whose project failed or succeeded within this phase
1.1	Analytical Tools / IT Infrastructure	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
1.2	Data	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
1.3	Employees Competencies/ Skills/ Knowledge	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
1.4	Organizational/Managerial Factors	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
2	**Prototyping**	All resources/capabilities which are either especially important / success factors for the prototyping or in which governments are lagging/ have difficulties (also considered as indicators for critical factors)	Especially attention to the interviewees whose project failed or succeeded within this phase
2.1	Analytical Tools / IT Infrastructure	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
2.2	Data	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
2.3	Employees Competencies/ Skills/ Knowledge	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
3.4	Organizational/Managerial Factors	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
3	**Implementation**	All resources/sapabilities which are either especially important / success factors for the implementation or in which governments are lagging/ have difficulties (also considered as indicators for critical factors)	Especially attention to the interviewees whose project failed or succeeded within this phase
3.1	Analytical Tools / IT Infrastructure	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
3.2	Data	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
3.3.	Employees Competencies/ Skills/ Knowledge	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
3.4	Organizational/Managerial Factors	Investigation of all resources/capabilities based on the categorization stated in chapter 2.2 + reporting of additional critical factors	
4	**Influence of DSSG**	Investigation of the assistance by the DSSG; Resources/capabilities the DSSG provided throughout the project process	
5	**General Impact on the Resource Base / Competencies**	Investigation of the impact on the organization while executing the data science project; Reporting of impact for which resources and capabilities	

Table 5: Analysis – Coding of Interviews (Excerpts)

Meaningful Units/ Representative Quotations /Transcript	Codes	Category	Theme
"In terms of infrastructure is not great for this kind of work. We are not allowed to do different things, and we are very much behind the times, and that is an impediment. We are behind the times with the variety of infrastructure. There are different security measures within the network (because it is a governmental network)." R1	**Lagging / Importance:** Tool Sophistication	1.1	**1**
"Further, breaking down barriers would help. Not just the GDPR but also clarify ethical concerns, facilitate an easier automated decision-making process, etc." R1 *"The privacy of the individuals is also important where you are driving the data from. One concern is always that when we engage a client in our program, we make it clear to that individual that there is going to be a substantial need for data collection from and over time. But we also try to understand what the purpose and extent of the usage. The clients need to understand that much of the privacy will be protected and that the use of that data will simply be to inform program operations – that they identify will be protected (e.g., for reporting, etc.) it is usually at an aggregate level. We try to make it as ensuring as possible." R2*	**Lagging / Importance:** Use Policy	1.2	
"The data opening issue is one of the main factors why we failed our project. The departments are nervous about sharing data, and the General Data Protection Regulation (GDPR) also amplifies this and lead to uncertainties" R1 *"Within the scoping phase, we failed because of data sharing and the infrastructure and how we secure share data. There is a need for improvement." R1* *"Governments are really good at collecting data. Part of the so valiance authority is to know what's going on in the environment and community. So, there is an information system (e.g., public health informatics, that collects a lot of data). The problem is that the data is often siloed. You have to go to a lot of different projects to reach it when it is collected. The data opening issue is one of the main factors in the first place. It is not in one specific database to look at in the first place. There are many databases, and they are on the federal, state and local level" R4* *"So, it takes a long time to assess the requests, to share data, and again- they are very concerned that they would not be able to share it and that the commission would veto the sharing of the data. So that is really a big constraint in what concerns data." R5*	**Lagging / Importance:** Accessibility /Sharing		
"We have to be very transparent, what we are doing with their data, which decisions we are taking, how we are making those decisions. So, I think it is that black box aspect of some of these algorithms that when we give them to the public, we need to explain them. So, we need to be super careful, that within data science and especially when talking about ethical concerns, everything is going to be explainable, open and transparent." R1 *"The opening process of data is also a topic at the beginning– Chicago is really good with being transparent with their open data port but there are certain sets which you just can't open within our departments. (e.g., personal health data). That will never be open to the public, because the way the model works you need to know where to do it (e.g., which house /address – within the led poisoning case) – so, it will never be part of an open government try of data set. But other things that contributed to its success are open data sets. So, it is needed to help – it depends on the projects. But in the long run, its good for governments to release the allowable data sets" R4*	**Lagging / Importance:** Opening Process		
"One adjustment is collecting more information to start a project. It is kind of surprising how few organizations collect intervention information if they do it or even focus on research." R6	**Lagging / Importance:** Bigness of Data		
"I think they should have high-quality data because I think more or less the major operating systems we already have in place in our areas. E.g., if you want to work with medical prescriptions, we have all that data already. So, I think we have a lot of good quality data within public administration. That means, however, a very big concern from those institutions because they have no to comply with the GDPR right from the beginning (so all the restrictions there are with opening the data, etc.) So, I reckon they are afraid to open to other organizations in the first place, e.g., because there might be problems about leaking data to the public and so on, & they are really concern that this GDPR would make them liable for using that data. So, I think we have really good quality data ...but it is very difficult to make it available. It is not just GDPR, we also have a national commission for data privacy protection, it is a commission for access to administrative documents, which, for instance, if	**Lagging / Importance:** Opening Process Quality Use Policy		

one institution wants to share some data for operational purposes which are coming in place when implemented, they have to ask this commission." R5			
"Well, what we noticed is when we started these partnerships, from the public administration side of the projects, their most concern was about data privacy/secrecy and access." R5	**Lagging / Importance:** Opening Process Use Policy		
"My impression is that technology and especially data science is already used widely. I think it is a matter of the extend – how sophisticated and helpful and how dynamic and quick it can be. There is a difference between working at the beginning with dozens and dozens of sources (a lot of work). The governmental partner institutions are often stretched." R3	**Lagging:** Source (Abundance)		
"So, a major point is understanding data policies in the first place, and it is not just us understanding them but also the organizations (external). They did not understand it- so we needed to look at contracts to see how governs over this data, what are the rules to using this data. It is just better governance overall." R4	**Lagging / Importance:** Governance Use Policy		
"Most of your partners really struggle. Because they have collected data for a certain purpose and not for our purpose and we need historical records. And we run into a lot of problems when we go to organizations, and we need the most fundamental thing that we need – to you have any data ', nor we do not.' One example we partnered with a police department and asked for data on that, and they said: they do not. One example, we partnered with a police department, they gave us data and in that particular case we are trying to predict whether an individual office is going to do something bad, meaning that we need individual officer information and what they gave us were image PDF's telling us how many times the department has a whole used force in the last quarter. Something like this happens the whole time." R6	**Lagging / Importance:** Integration		
"Even if our people have (basic) skills in coding, we probably need more in terms of IT skills, program all that staff." R1	**Lagging / Importance:** Technical Skills Analytical Skills	1.3	
"One example, we partnered with a police department, they gave us data and in that particular case we are trying to predict whether an individual office is going to do something bad, meaning that we need individual officer information and what they gave us were image PDF's telling us how many times the department has a whole used force in the last quarter. Something like this happens the whole time. In that case, technical knowledge was missing because they did not even know what system they are using. We wanted to help them and asked at the organization how to get data out of the system, and they said they were not their clients. And the police department said 'yes, we are.' We had a week-long conversation between the two of them where we determined that they were not using the software. They did not even know what system they use." R6	**Lagging / Importance:** Technical Skills		
"The issue is on the resources and skills. They need to think of a way to put in place those projects. It should be someone dedicated to data science and data management and to help them in the scoping phase. They need to have someone who is not totally focused on daily operational tasks. It is not that easy to do it – since there are so many constraints. But someone who can see the potential outcome of the data, there needs to be someone who cares for data, etc. They have to have a data protection officer because of the GDPR. I think they would think about data manager, chief data officer (e.g.)." R5	**Lagging / Importance:** Analytical Skills		
"Data Science is not at the forefront of people's priorities. There is not a huge emphasize on data science, but I think people within the department really just don't know what possible and they don't take much time to engage with it, so I think if they knew what can be done (this was also one reason for this project, trying to demonstrate the art of possibilities of the speed)." R1 *"The main idea to start the project was the idea that hopefully there will be awareness raised for this kind of work." R1* *"Scoping at the beginning of the process requires them to think about possible interventions that they can take and also the ability to think about how to implement it later on if the project is successful." R6*	**Lagging / Importance:** Organizational / Managerial Factors Sense for Economic Potential/Awareness	1.4	
"There is a branch within this department which is engaged with the private sector - called innovation and specialization branch (department). But within our department, there is not much going on with data science." R1 *"One adjustment is collecting more information to start a project. It is kind of surprising how few organizations collect intervention information if they do it or even focus on research." R6*	**Lagging:** R&D		
"Further, breaking down barriers would help. Not just the GDPR but also clarify ethical concerns, facilitate an easier automated decision-making process, etc." R1	**Lagging / Importance:** Ethical		

"There is a further certain manager who is or are not keep about the data science structure, so it is not like in businesses wherein a tech company are all keen, these are different people who might not be keen about doing data science." R1 *"Automated decision making is coming with a lot of ethical concerns. I think it has to be some kind of human interaction at some point. Especially for governmental institutions because there is a perception rightly or wrongly government is interfering and there a conspiracy at work. We have to be very transparent, what we are doing with their data, which decisions we are taking, how we are making those decisions. So, I think it is that black box aspect of some of these algorithms that when we give them to the public, we need to explain them. So, we need to be super careful, that within data science and especially when talking about ethical concerns, everything is going to be explainable, open and transparent." R1*	Decision Support		
"You need to be able to show the policymaker, e.g., that you are able to make a difference and that you can serve the society better with algorithms. There are further certain managers who are or are not keep about the data science structure, so it is not like in businesses wherein a tech company are all keen, these are different people who might not be keen about doing data science. In terms of managerial skills, there is constraint like legal and managers are super risk averse. If you take legal advice, the would rather say no to specific ideas." R1 *"It is also the leadership of an institution. Before you let the DSSG team work on a specific issue, there are often 2-3 people at the partnering side who need to do something, and it just happens more slowly." R3*	**Lagging / Importance:** Leadership		
"The linkage problem is also really huge. The people have a problem sharing data between the DSSG or other departments, organizations – there is a perception of risk there. A legal way to share data and a willingness to share data is also pretty important – but sometimes the departments are really risk averse and everything that might go wrong, are negatively affect their reputation and rights. So, the minute we told them to share data with an outside party, it is getting really difficult!" R1	**Lagging / Importance:** Alignment IT/ Data + Business Politicians/Politics/Legal		
"Especially when you work with different politicians where certain ideologies are in place and then you can always get a policy through based on evidence. Northern Ireland in particular – we have not had a government here for several years – we have a local assembly who are responsible for policy-making on a local basis. Civil services are mostly on their own, that again had an impact since we are making decisions, we are sometimes not able to do – which sometimes might require a minister to sign of. The minister part and policy part are usually coming in place in a later part as the scoping when there are more resources committed to it. We have in the scoping phase some licenses to work with research bodies. If we were in the implementation phase, we would need some kind of sign off, and here in the absence of a minister, that is a legal question which is also quite hard to tackle." R1 *"Reconfiguring resources and being dynamic in the changing landscape is difficult, as governmental institutions are, e.g., locked in procurement regulations." R1* *"Regulations are especially important in the scoping phase. We did not see all the problems that were coming in the scoping phase because of the regulations" R3* *"So, it takes a long time to assess the requests, to share data, and again- they are very concerned that they would not be able to share it and that the commission would veto the sharing of the data. So that is really a big constraint in what concerns data." R5*	**Lagging / Importance:** Politicians/Politics/Legal		
"By the time you managed to do one change, things moved on already – we are always playing catch up. That is again one reason while partnering with an outside organization like the DSSG is an advantage since we are getting access to people with the right skills, are used to have the right tools." R1	**Importance:** Collaborations / Partnerships Time Management		
"The interaction with the human and the machines should be enhanced, but we would like to have human elements and their decision making still within the projects." R1	**Lagging / Importance:** Alignment Resources		
"But for sure, the technical capabilities to be a little bit more specific, instead of relying on outside knowledge to be able to correctly frame a problem to be properly solved. That is also why we started to make these partnerships with the research community because we were a little bit afraid that if we made this partnership with enterprises/companies, they would sell a solution to the public administration, which perhaps wasn't exactly what they needed. And that research community would be more a proper advisor than just a seller of solutions. So, I think when you're talking about scoping, I think that is perhaps the main problem. They lack the skills to exactly define what can be done, what data is needed and how can it be done." R5	**Lagging / Importance:** Collaborations / Partnerships Managerial IT		
"Also understanding all the policies form governments of internal and external data sets. If we knew beforehand, it would have been faster and quicker to negotiate, or these delays occur because we are using data from the state or federal government.	**Lagging / Importance:**		

So, a major point is understanding data policies, and it is not just us understanding them but also the organizations (external)." R4	Politicians/Politics/Legal Time Management		
"During a prototyping phase, the departments have to live with the contains of the IT infrastructure." R1 *"So, there was a hand-off that was incomplete at that time, but it gives us some ideas of what might be some of the predictive elements that will be in the model. What we did as a follow-up, was to pick up where they stopped. We did not have the infrastructure and skills to follow the machine learning approach- so we went back to regression and modeling approaches. And approximately 1 year ago, we completed a report on that and now we are working with a third party who is going to help us develop some algorithms as far as embedding an application on their database that can reflect when clients data are on high risk, and it will be part of our operations reporting so that a nurse could be better attempt to it." R2* *"In the prototyping phase, many departments and people are getting involved. The technical infrastructure is crucial, and stakeholders need to be convinced that the project is useful. Further, the coordination between the departments needs to be enhanced." R3* *"We have an internal probability infrastructure which technically would allow sharing the data when it's needed for the prototype, but there are also some challenges from the technical side – and we have in our national statistics office, they already collect data from several sources for statistical purposes, and they have already in place an infrastructure to make some of that data available to researchers under anonymous and secure conditions – but still, on a technical level, they even say they have some difficulties to get the data out of some institutions. And I think that is perhaps because those institutions are afraid, they would infringe some regulations about privacy and sharing data, etc. That is one of the major obstacles." R5*	**Lagging:** Analytical Tools/ IT Infrastructure	2.1	
"Often the data are not really easily accessible, complicated to get access to it which is especially a burden within the prototyping phase." R3 *"In my case, it was inaccessible silo data as a result of the reject these were brought together to a central source, and now, they can work with it. And maybe for the future, they will have an easier time to integrate the data that they need." R3* *"Yes, when we presented the question, really at the beginning of the second year and they decided that they will work with us that summer and assign a team to us, I think at that point in time, I wasn't aware of what the capabilities or the complexity of the project will be that summer. And clearly what the data requirements will go to be and what the time requirements will go to be for the team. So, we could have prepared a little bit more, maybe also come into the summer fellowship better prepared for what they are going to need from us. What are they going to need to have in place for the first week of the fellowship etc. to really get this could and to use their time as effectively as possible? Because what I think, to some extent, they started on the project but had to wait around a little bit – as far as getting the full datasets together. That would have helped if we would have had some more awareness on our ends." R2*	**Lagging / Importance:** Accessibility/Sharing		**2**
"So, that year the team actually did a Machine Learning approach to trying to address this question and what happened was that we quickly learned that they needed a lot more data than we initially anticipated. We obviously spent a long time pulling datasets, doing a little bit of cleaning and transferring the data, etc. and we started to generate some results from the modeling but literally ran out of time by the end of the summer session." R2	**Lagging / Importance:** Bigness Integration	2.2	
"The GDPR who was coming this summer was also an obstacle to overcome, especially in the prototyping phase. Maybe in the implementation phase as well, but in general I guess there you already have access to all data, and you managed it properly, in a secure way or anonymize it or whatever." R3 *"In the US there are other policies than the GDPR in Europa in place: we have different rules of laws like HIPPA etc. – but those do not stop the implementation. They do protect people which is good." R4* *"For instance, let me go off track a little bit – there was one of the most conservative organizations that we have, most focused on operational results. And I knew there were a lot of research institutions who already had an approach, to use that data because they have very high-quality data (big data) on spending, to make studies and prototypes – and they always said no. Within this initiative, we were able to convince to participate with three projects. I had to meet them several times, and they were always concerned about this privacy-issues, and they wanted to see if we release data and if something happens, if citizens no longer trust us, because it is such highly-sensitive data. And we managed to convince them that for the first time, they would do a partnership with the research community to prototype and to seek*	**Lagging / Importance:** Use Policy		

projects within this area. So, I think this definitively has an impact on even the most conservative organizations within the public area (sensitive to realizing data, etc.), they are now more open to this because we have started this initiative." R5			
"The data/analytical capability from the project I worked on worked quite good. Most of the capability on the data science part was done by the DSSG team, but there was still technical work to do to make this work of the DSSG available. E.g., anonymizing it before the DSSG can see it (because that is required by law), using SQL, but just from a technical perspective – there are other parts that need to be considered in the prototyping phase." R3	**Lagging / Importance:** Technical Skills	2.3	
"So, there was a hand-off that was incomplete at that time, but it gives us some ideas of what might be some of the predictive elements that will be in the model. What we did as a follow-up, was to pick up where they stopped. We did not have the infrastructure and skills to follow the machine learning approach- so we went back to regression and modeling approaches." R2 *"One of the things what we can do better to enhance the process of data science projects itself is to train some of our people for the opportunities that come with data and unlocking the skillsets, understanding data and using it on a day to day basis." R2* *"I have seen unrelated projects that there is a willingness to learn from the IT/analysis/tech team of the institution (e.g., with workshops) which is especially important for the implementation phase in data science projects. But it should happen before the implementation with the purpose of implementation. Because you need to have these abilities all ready to make a decision." R3* *"We ended up having another partner for the process, but he was not the right on. He was an academic who was interested in what happened in the past but what we really should have done is build a tool that can be used in the future." R6*	**Lagging / Importance:** Employees Capabilities		
"Some people (from the DSSG side) also think that it would be helpful if some people only have this project in mind and would push it further at the partnering institution. I would call it the ability to focus since this is additional work for them. So that there is always someone how is the counterpart of the DSSG whenever they need a contact person, but then you also need some director who gives the directions and is a decision maker within the project. E.g., in my project, it would be doctors who would say what makes sense. Those domain knowledge persons would be a helpful resource to get involved and not just the technical persons. Someone who would push it forward in a dedicated manner beyond their own part, taking over the big picture of the project." R3	**Importance:** Domain Knowledge		
"The city has two data scientist, and they are not in our department (department of innovation). So, we are reliant on internal partners. We have epidemiologists that are well trained in statistics but not in data science. The workforce needs to be trained in data science methods, and that would be a upskill in the future. Within the scoping – we can do it and build up a hypothesis. But do the actual role in prototyping, and later on, we would need data scientists. We learn a lot, but we need support in general." R4	**Lagging / Importance:** Analytical Skills Domain Knowledge Technical Skills		
"Yes, I think in some respects we are quite capable. What we have learned is that we have some skill sets that are little limited at the moment. Particularly as far as using Machine Learning approaches and we recognized the potential of our database to use that methodology. I think the next big step for us is probably to bring in that skill set into our operations so that we could better use that approach. That said, we are pretty well staffed in other aspects (regression modeling and other types of analytics processes). And we recognized a need for probably for the next level – just bringing in talent and some experiences that could take us into Machine Learning approaches. So, I think that is next for us." R2	**Lagging / Importance:** Analytical Skills		
"And the Human resources are also a problem – but more in the later phase (e.g., Prototyping). Data science projects are usually not their main task to fulfill. It is usually the first barrier because even yourself often don't really have any time to do it. Priority is the society, and the data science is a lower priority which you can do in the background. For employees, this usually means its added to their everyday job." R1 *"In the prototyping phase, it is definitely an obstacle that other work is going on. Even if there is enthusiasm, there need to be someone to push it forward. There are several people involved in different departments, and it is not they do not want, but even if they do want, they have their regular work and maybe several projects at the same time. Mostly it is just extra work in addition to what they are already doing. That is the biggest obstacle." R3* *"Mostly because they need to stick to their regular work and are used to do it and there is a lot of it. If you want to change the governmental institutions towards a data science approach in a very ambitious way, then there is a need for people focusing on the project during the development as they are often buried with regular report work. That is why they think they do not need this advanced stuff." R3*	**Lagging / Importance:** Peoples Resources	2.4	

"Within one project, there was a guy whom we talked to, and just one week later he said I am not working here anymore, and you need to talk to someone else, and the person also said that it is his last day and there was no one else to talk about this. So, there wasn't someone to talk to, and we ended up not having a partner anymore during the prototyping process." R6			
"There is outside help needed. That is why in our initiative we financed 19 projects to either prototype or implement, and we asked specifically that at least one of them would be in a partnership between one organization within the public administration and at least one research center from outside. We left companies out of this financing mechanism, which was from the ministry of science, and they could not find our companies. We did that because we knew that these organizations within the government systems need, e.g., knowledge and everything else from outside the public administration. We think that it was better to start this partnership with the scientific community than just leaving this governmental organization to subcontract services from private companies. Which is more or less how they operate when they need to develop some kind of technological system." R5 *"I think most of these organizations do not think in terms of prototyping. When they think of launching the project, they mostly think about the final product. In some parts, they are changing, but I think most of them do not have the financial resources to spend things on prototyping and try-out-periods. This is something we are trying to change within our initiative, was to say 'ok, this is to – the partnerships are between the public administration and the scientific community – and the role of the community is not to deliver an end product. You maybe should consider this as prototyping a project. But they have no problem with that because they did not finance any of it – it was all financed by the science party." R5* *"And we managed to convince them that for the first time, they would do a partnership with the research community to prototype and to seek projects within this area. So, I think this definitively has an impact on even the most conservative organizations within the public area (sensitive to realizing data, etc.), they are now more open to this because we have started this initiative." R5*	**Lagging / Importance:** Collaborations / Partnerships Funding		
"Yes, I totally agree with you. There is a positive impact on the organization while executing a data science project. That is something that I can say now that, that this initiatives show is that it opens a lot of minds within public administration to the importance of data science and prototyping. For instance, let me go off track a little bit – there was one of the most conservative organizations that we have, most focused on operational results. And I knew there were a lot of research institutions who already had an approach, to use that data because they have very high-quality data (big data) on spending, to make studies and prototypes – and they always said no. Within this initiative, we were able to convince to participate with three projects. I had to meet them several times, and they were always concerned about this privacy-issues, and they wanted to see if we release data and if something happens, if citizens no longer trust us, because it is such highly-sensitive data. And we managed to convince them that for the first time, they would do a partnership with the research community to prototype and to seek projects within this area. So, I think this definitively has an impact on even the most conservative organizations within the public area (sensitive to realizing data, etc.), they are now more open to this because we have started this initiative." R5	**Lagging / Importance:** Collaborations / Partnerships		
"I do think that be biggest problems are ultimately people's problems for the prototype but also later on. Technical there are issues but more secondary in contrast to the characters that are involved. Again, and again, when it comes to governments, we run into problems where an election comes up, or someone is fired and then suddenly there is no more support for the project. That is what happened to one of our past projects. They had recent elections, a conservative government comes in, and they do not want to give credits the previous government is doing. As a result, the project will probably fail in implementation because of this problem (here: Chile project). Since the new government no longer has an interest in supporting old decisions. *In business, they do not change the idea across people. But when politics are involved, the success of the former persons is a problem. Along with that in the US, we run into problems, that it can be very difficult for governments to get rid of its people compared to companies. They can be people who say (here in a public health institution) I am wasting my time with this job, but some governments are not able to do something about it." R6* *"So, we have several projects, but it is a lot within the prototyping phase. A lot has been driven by Brexit with the need of new statistics and new analysis we have not been done before – so there is for sure demand.* *We do have some projects implemented. E.g., Brexit problem. Like Northern Ireland, we have a border with the Republic of Ireland. And we did a project with data science how many crossings at the borders are there." R1*	**Lagging / Importance:** Peoples Resources Politicians/Politics/Legal		
"In the prototyping phase many departments and many people are getting involved – there is some technical stuff which is missing, sometimes convincing people that	**Lagging / Importance:**		

the project is useful within other departments also like coordination and control between the departments" R3	Peoples Resources Sense for Economic Potential/Awareness		
"But in public health, where we look at prevailing practices, yes, it is the beginning of transformation. They just don't have the tools and skillsets to finish or start." R4	**Lagging / Importance:** Tool Sophistication	3.1	3
"Their lack of the resource of skills, they also lack tools, and they also lack the ability to get developed projects in place. The data or the systems that store the data is not really prepared for analytical purposes. They were built more or less specifically for operational purposes, but they also need financial and technical resources to take full advantage of that data." R5	**Lagging:** Analytical Tools/IT Infrastructure		
"One more is that governments often are having technical requirements that businesses do not have for example: having to build websites which are much more accessible to a wider variety of browsers. Governments need to worry about that. I know that in the US, there are certain requirements for governments." R6	**Lagging / Importance:** Additional Technical Requirements		
"I think that culture and the history, as well as regulations, collaboration, and legacy systems, are reasons why governments are lagging behind since this is usually burdened for governments to implement their projects in order to be agile (e.g., one reason is funding)." R6	**Lagging:** Legacy Systems		
"In the US there are other policies than the GDPR in Europa in place: we have different rules of laws like HIPPA etc. – but those do not stop the implementation. They do protect people which is good." R4 *"So, I reckon they are afraid to open to other organizations in the first place, e.g., because there might be problems about leaking data to the public and so on, & they are really concern that this GDPR would make them liable for using that data. So, I think we have really good quality data …but it is very difficult to make it available. It is not just GDPR, we also have a national commission for data privacy protection, it is a commission for access to administrative documents, which, for instance, if one institution wants to share some data for operational purposes which are coming in place when implemented, they have to ask this commission." R5*	**Lagging / Importance:** Use policy	3.2	
"If we could better connect the results internally and make the internal data sharing easier (within the departments) that would be good." R4	**Lagging / Importance:** Accessibility/Sharing		
"The enhancement of the skills of the people is not necessarily needed to the scoping phase, but while you implemented a project, the people need to be capable to maintain, retrain the model to make sure that is actually reflecting changes in the data. So, you do need the skills to keep it relevant. Within our organization, we have people here who are capable of these IT skills and can code with python or R but again the organization needs to invest in more resources to give people more time, to develop their own skills. We could also recruit people from outside who are capable of doing this, the problem would be that you are paying an extra for these people and since we have a strict wage structure within the public sector. My preference would be (since we are capable of doing a lot of this stuff by ourselves) to upskill." R1 *"I think it is a different story in the implementation phase. Now it is all writing in python like the modeling and dashboard etc. the output and if someone needs to modify it is is going to be a challenge, because the analytics people have done something different in a different matter." R3* *"I have seen unrelated projects that there is a willingness to learn from the IT/analysis/tech team of the institution (e.g., with workshops) which is especially important for the implementation phase in data science projects. But it should happen before the implementation with the purpose of implementation. Because you need to have these abilities all ready to make a decision." R3* *"I think this depends. When you look at public health. To say that adoption of data science is generally excepted – I would say yes. But looking at how these organizations are set up, they do not have the skill set, they are generally lagging. I do not think it is a negative outlook for the opinion of data science, but I think the skill sets and the resources available like funding to get the specialist skill set into public health are missing. So, in one sense it is favorable but not in all, e.g., in emerging (I don't know if we can consider this emerging, as if we compare it to Netflix to Google and Amazon, they have it for years) – but in public health, where we look at prevailing practices, yes, it's the beginning of transformation. They just don't have the tools and skillsets to finish or start." R4*	**Lagging / Importance:** Analytical Skills Technical Skills Employees Capabilities	3.3	
"They lack the resource of skills, they also lack tools, and they also lack the ability to get developed projects in place. The data or the systems that store the data is not really prepared for analytical purposes. They were built more or less specifically	**Lagging:**		

for operational purposes, but they also need financial and technical resources to take full advantage of that data. So, they are aware of the importance, we have been contacted - we had more than 50 applicants for this financing project, but they want to do it. They are aware of the importance, but they also recognize that they lack certain resources – they are pretty much focused on operational reasons." R5 *"And if they have enough resources to implement, they are very much more focused on daily operational routines. In public organizations, it is rather difficult to create the kind of skills which provide governments with the opportunity to participate, maintain and continue with data science projects." R5* *"I think it would be difficult, to view that some organizational structures would change, because of these projects. At least at this point. For one, it is difficult to get the right skills in public administration for maintaining these projects, I think they are becoming more aware – perhaps, one or two would be willing to hire some data scientists/chief data officer, they would perhaps improve a little bit their data structures and would improve the access mechanisms, but right now – in my opinion – they will be no major changes within the organizational structure because of this. I think one or two would create a small innovation hub/center/ team – but even then, I would think it would be very difficult because there are a lot of constraints on hiring people regarding financial resources, etc." R5* *"Database administrators could be helpful to implement, people who can learn under staff has proven to be useful. Even if some partners have good people, so many don't even know what we are doing. And the partners who participated twice within the fellowship showed the ability to learn (e.g., how to build a model) as they were able to do it on their own from there on." R6*	Employees Capabilities		
"A large portion of our organizations are nurses who work in the field and provide educational components to clients. They are skilled from the nursing thing, but it is probably a mixed bag to use data to inform what they are doing to recognize the utilities as far as how analysis might enhance their practices after implemented." R2	**Importance:** Domain Knowledge		
"Had I to do the project over again, I might have either sharped it around at other Universities (e.g., University of Colorado) to see what was available in the computer science program up there. To see if there's someone who can continue with the machine learning part of it." R2	**Importance / Lagging:** Analytical Skills		
"The city has two data scientist, and they are not in our department (department of innovation). So, we are reliant on internal partners. We have epidemiologists that are well trained in statistics but not in data science. The workforce needs to be trained in data science methods, and that would be a upskill in the future. Within the scoping – we can do it and build up a hypothesis. But do the actual role in prototyping, and later on, we would need data scientists. We learn a lot, but we need support in general." R4	**Lagging / Importance:** Analytical Skills Domain Knowledge Technical Skills		
"Especially when you work with different politicians where certain ideologies are in place and then you can always get a policy through based on evidence. Northern Ireland in particular – we have not had a government here for several years – we have a local assembly who are responsible for policy-making on a local basis. Civil services are mostly on their own, that again had an impact since we are making decisions, we are sometimes not able to do – which sometimes might require a minister to sign of. The minister part and policy part are usually coming in place in a later part as the scoping when there are more resources committed to it. If we where in the implementation phase we would need some kind of sign off and here in the absence of a minister, that is a legal question which is also quite hard to tackle." R1 *"We could also recruit people from outside who are capable of doing this, the problem would be that you are paying an extra for these people and since we have a strict wage structure within the public sector. My preference would be (since we are capable of doing a lot of this stuff by ourselves) to upskill." R1* *"But even securing the data appropriately, it needs to be done really properly in the last process - you need to have a lawyer who checks if everything is ok, you need to have the people who actually have the databases, (maybe there are even not the analyst you are working with during the DSSG time)." R3* *"As a success within the implementation. But in general, we look at the projects more like learning. There are a lot of policies throughout all departments, which we needed to bypass when implementing the project (e.g., wage structure)." R4* *"So far, we were lucky. No one rejected our projects with funding so far. But let's see what will happen after the election in February 2019. There is a new election." R4* *"I think that culture and the history, as well as regulations, collaboration, and legacy systems, are reasons why governments are lagging behind since this is*	**Lagging / Importance:** Politicians/Politics/Legal	3.4	

usually burdened for governments to implement their projects in order to be agile (e.g., one reason is funding)." R6			
"The regulations we have can be seen differently when we talk about later phases. We have a compliance officer in place, legal counsel, hippa compliance officer… etc. so we need to have a substantial amount of resources. We have a chief information officer who runs our data systems, hippa compliance, firewall requirements, etc. and for all those complaints to ensure the standards. So, we are quite aware of it. I would not say that it is an unnecessary burden – we respect the intensity of these rules. But we also know that you have to have a lot of resources to stay compliant with the rules." R2	**Importance:** Politicians/Politics/Legal Peoples Resources		
"My impression is that technology and especially data science is already used widely. I think it is a matter of the extend – how sophisticated and helpful and how dynamic and quick it can be. It is different when working at the beginning with dozens and dozens of sources (a lot of work). The governmental partner institutions are often stretched. Then, there is a difference at the end of the process between putting together a report at the end of the year and making something that is providing information (as things are happening). The importance is high, it is just not leveraged to a full extent." R3 *"The thing is that – maybe the public institutions often have a light use of data, but it's not leveraged to the full extent – after implementation it's often done for reports for foreign policy and its really useful but often analytical tools are not used, e.g., every week, month in some position to guide decisions in real time. If you want to have the end of the year reports that are useful, that are crucial for next year's policies. – I think they have the tools for that, but they will not be able to build a model that has an output every week and changes the output depending on how things are changing." R3* *"I think we have recognized the potential - we always collected a lot of data, that was just the nature of the program – but we initially only used it later on for the purpose of reporting, not for analysis to drive program operations or to understand the nature of the operations better. Do the type of deep-down analytics that tells you what's influencing your outcome from an analytic level. We progressed in this direction in the last eight years and developed resources for that. When I came on board, we were relatively week in our awareness and capabilities, and we progressively build out. It will continue to grow." R2*	**Lagging / Importance:** Sense for Economic Potential/Awareness Decision Support		
"When implemented, it is much easier for businesses to stretch and align everything between departments and to have data science as a base for all decisions." R3	**Lagging:** Alignment Resources		
"So, there is some education that has to go up to the leadership level to build awareness – so that 'if we have skillset of that nature, here is the potential that we can achieve with that'- recognizing that this also is a salary grade that might be a little bit higher than what we previously anticipated in comparison to those people we usually recruit. So, I think the burden or challenge is simply building the awareness about what those types of individuals can do for the organization and correspondingly what the organization would need to invest to bring that capability on board when it comes to implementation." R2	**Lagging / Importance:** Leadership Sense for Economic Potential / Awareness Funding		
"The organizational structure, in particular, is a think since data is not based in all departments and since several departments are involved in data science projects, it is quite hard to overcome the obstacles. It is also a kind of silo using of the data in one specific project and is therefore often not implemented in other departments. Definitely, in other departments, some might not realize how useful these data science projects can be because they have not been used to inform their decisions or build their strategies based on data. Many of these departments not even at all." R3	**Lagging / Importance:** Sense for Economic Potential/Awareness Decision Support Communication		
"In the implementation phase not just pushing forward but sometimes asking how the project can be useful to the different users/stakeholders in different departments but also convincing others that it is useful within the organization. Some of them will realize it directly what we should aim for, but maybe there are others who need to be on board as well, and maybe they are not used to think about data in this way. Someone needs to talk and explain them." R3	**Lagging / Importance:** Sense for Economic Potential/Awareness Communication		
"I would not say all of them know about the importance of data science. All the ministers which I work, they are very very keen on this topic and very aware of the importance of data science for public administration when it comes to implementation." R5 *"A large portion of our organizations are nurses who work in the field and provide educational components to clients. They are skilled from the nursing thing, but it is probably a mixed bag to use data to inform what they are doing to recognize the utilities as far as how analysis might enhance their practices after implemented." R2*	**Importance:** Sense for Economic Potential/Awareness		
"In the implementation phase, the technical involvement also needs to grow. Here in my project, the technical part was fine, but if there will be finetuning and changing a lot of DSSG work, then we would need more technical involvement. So, technical knowledge and technical management will grow and the managerial part	**Lagging / Importance:** Managerial IT		

might drop a little in its importance but will also be a huge part within the implementation phase." R3 *"So, I think, starting from scoping - I think managers from public organizations, have a general idea of some problems they think data science (AI) could help them solve. However, missing the technical knowledge to see exactly, which data is needed, how can they get it and prepare it - and which techniques are available to address those problems." R5*			
"Even if the team leader wants to implement it and change the system, there is someone above that person, and someone above that person, leading to the difficulties of governmental institutions." R3	**Lagging / Importance:** Decision Support Politicians/Politics/Legal		
"Had I to do the project over again, I might have either sharped it around at other Universities (e.g., University of Colorado) to see what was available in the computer science program up there. To see if there's someone who can continue with the machine learning part of it. *Since then, I identified local resources which potentially would have the capabilities and interest to do something like that although it took 1-2 year to identify it and we were gone in a different direction with the project on that point. There are things we could have done before the project launch and things as a followup to make it more efficient. Lessons learned!" R2*	**Lagging / Importance:** Collaborations / Partnerships Time Management Peoples Resources		
"It is hard to measure when you say how much – because of funding wise when projects are in place not as much but in the investment of time and my resources probably more than other public organizations." R4 *"Their lack of the resource of skills, they also lack tools, and they also lack the ability to get developed projects in place. The data or the systems that store the data is not really prepared for analytical purposes. They were built more or less specifically for operational purposes, but they also need financial and technical resources to take full advantage of that data. So they are aware of the importance, we have been contacted - we had more than 50 applicants for this financing project, but they want to do it. They are aware of the importance, but they also recognize that they lack certain resources – they are pretty much focused on operational reasons." R5* *"I think one or two would create a small innovation hub/center/ team – but even then, I would think it would be very difficult because there are a lot of constraints on hiring people regarding financial resources, etc." R5*	**Lagging / Importance:** Funding		
"It is multifactorial. It is not one thing – it is usually several things. When you look for a perfect storm of success, we do not look at if we had the money, we will be ok- because of the infrastructure, the environment to handle these advanced analytics and all player apart. I would use if you use your mechanics, how things are set up if you have departments which are relying on that grant funding. The grants and the organizations like governments that a wart these grants – by default, this creates a silo. Like: this money should be used for this database or only this project. Anticipating that a project comes on, but we do not have the budget for it (e.g., we did not allocate for it, people did not plan for it.) Now we try to look ahead and say ok – here are the people that can work with our partners. How can we extend our capacity with partnership – that is really key. How can we check in advance in future and apply to allocate funding for the project we perceive will be presumed? It takes planning. When it was coming out like this, we did not foresee funding and partnerships. The timing was right. Now it is less ad hock, it is more internalized and therewith more purposeful. I think we plan for it, allocate, we commit resources. It is a start." R4 *"Our projects are currently really institutionalized since we want to finance partnership with the public administration and research organizations. (Several projects)" R5*	**Lagging / Importance:** Alignment Resources Collaborations / Partnerships Funding		
"I think the number one is the partnership. The partnership was not just University of Chicago or DSSG, but we have partnerships within the department in the led poisoning project, we had a top down and bottom up the type of collaboration within the department." R4	**Lagging / Importance:** Collaborations / Partnerships		
"There is outside help needed. That is why in our initiative we financed 19 projects to either prototype or implement, and we asked specifically that at least one of them would be in a partnership between one organization within the public administration and at least one research center from outside. We left companies out of this financing mechanism, which was from the ministry of science, and they could not find our companies. We did that because we knew that these organizations within the government systems need, e.g., knowledge and everything else from outside the public administration. We think that it was better to start this partnership with the scientific community than just leaving this governmental organization to	**Lagging / Importance:** Collaborations / Partnerships Funding		

subcontract services from private companies. Which is more or less how they operate when they need to develop some kind of technological system." R5 *"I think that culture and the history, as well as regulations, collaboration, and legacy systems, are reasons why governments are lagging behind since these are usually burdens for governments to implement their projects in order to be agile (e.g., one reason is funding). One more is that governments often are having technical requirements that businesses do not have for example: having to build websites which are much more accessible to a wider variety of browsers. Governments need to worry about that. I know that in the US, there are certain requirements for governments." R6*			
"And approximately 1 year ago, we completed a report on that and now we are working with a third party who is going to help us develop some algorithms as far as embedding an application on their database that can reflect when clients data are on high risk, and it will be part of our operations reporting so that a nurse could be better attempt to it." R2	**Lagging / Importance:** Collaborations / Partnerships Decision Support		
"So, the communication and with external departments- when you look at success its people first, before technical requirements." R4	**Lagging / Importance:** Communication		
"It is hard to have an overall picture of the whole administration because right up to now there was some software being developed within several administrations, but there wasn't a collective awareness that this was part of a collective effort let's say. So, when we started this, we got the impression that perhaps there will be some data science / AI projects within those organizations, but we were not aware of them, and I think no one was aware of that. But when I started contacting some of these organizations, I realized that there were already some projects being developed within them. But right apart from some projects that we have already started, and some other projects that we know that some of those originations are still interested in developing, I still don't have a clear picture of what we had before this. I know for instance that our national statistics office has some projects within this area - financed by the euro-stats and when I contacted them and asked which kind of role they would play in our initiative, they made a presentation, and they showed me several projects which were underway in this area. And I know that also some organizations might have some projects in this area, but it is apart from the national statistics office it is difficult to know them. So, it would be important to see how others did the implementation." R5	**Lagging / Importance:** Alignment Resources Communication		
"For ethical concerns, it is important that within the projects we are doing, there are no purity results. We are trying to estimate the risk of a population to provide a service. That is not tremendous to their home or the environment. If we use risks scored and we look at the top highest risks, I think we estimated incorrectly. The worst that happened there, they would get a home inspection which is negative for led – in that sense, how we order and priorities, using a predictive model to provide a service. With others, it might be different. E.g., Those with a purity outcome. – e.g., police models." R4 *"I think the ethical concern is important. I mean some people really think of data as being very sterile and not capturing the essence of all the things that are going on in an operation. I am sensitive to that topic. I always encourage a mixed method approach for our data science projects which have been implemented. So, we are not only given the representivity what data might present but also use things like focus groups or interviews, observational studies so that we get the color which fills in what the data generally describes." R2*	**Importance:** Ethical		
"From quality improvements in the results to understanding the benefit of collecting data and having the data analyzed. I think at the leadership level amongst our administrators (there is some variety) to allow them to think about how data can better inform their decision-making. It is a mix as using data to drive a decision and using data from the field or more qualitative types of information to drive decisions. There is a need to find the appropriate balance in terms of the data drive and qualitative stories that drive the way we make decisions. *There are some needs there to define the right balance across the departments within organizations. " R2*	**Lagging / Importance:** Sense for Economic Potential/Awareness Leadership Decision Support Alignment Resources		
"The ethical questions were only raised from the side of the research communities. Some of the research institutes said, 'okay we are going to implement this, but we need to address these ethical conditions.' On the other hand, we are now preparing a broad strategy for artificial intelligence in Portugal. Ethics are certainly one of the specific issues that would have to be addressed. And we are thinking about the best way to do this – and I think we would have to put in place and ethical AI-commission, to see whether some of the projects have really indications on the ethical side / have an automated decision-making-system. We have looked into the experience of the Canadian government, where they have an assessment tool, which is available online, to assess the risk of every data science artificial project. Their approach is, we will not consider every project, because there are some projects – which from the ethical sense, they do not have that many risks, we will focus on	**Lagging / Importance:** Ethical Collaborations / Partnerships		

those projects where the tool tells us that this is a risky project from the ethical point of view. There is a set of constraints. So, I think the ethical issues on these projects are very much concerned on a higher level. I think that most of the organizations still look at these projects purely from an operational point of view, and they think 'well, if we have, e.g., a system that rates the risk of someone not paying the full amount of taxes, okay – this is a technical system, the better the system, the better for us.', I think they are not well aware that there might be some ethical concerns of having an automated system in place, that selects economic targets for inspections." R5			
"In governments in general, I would say that institutionalizing innovation is hard in general. For governments, non-profits, etc. And that the fusion of innovation throughout that organization both tasks are hard." R4	**Lagging:** Alignment IT/Data + Business		
"Whereas in some years passed, it would be more of a pipe dream as far as being able to move in that direction to internalize projects relatively quickly." R2	**Lagging:** Time Management		
"I really think that the problems lie within the resources that they do not have and the pressure to deliver operationalized results, which will not give them enough time to dedicate their attention to other, less important things like switching to a different system in operational terms, which I think is the main problem." R5 *"And if they have enough resources to implement, they are very much more focused on daily operational routines. In public organizations, it is rather difficult to create the kind of skills which provide governments with the opportunity to participate, maintain and continue with data science projects." R5*	**Lagging:** Time Management Peoples Resources		
"Database administrators could be helpful to implement, people who can learn under staff has proven to be useful. Even if some partners have good people, so many don't even know what we are doing. And the partners who participated twice within the fellowship showed the ability to learn (e.g., how to build a model) as they were able to do it on their own from there on" R6	**Lagging / Importance:** Peoples Resources		
"There is always a change in administration just by the sign. So, when we create these innovations, I think the important part is to operationalize the innovation and ensure it sustains. Unfortunately, we are driven by grants. But once you can operationalize a project, you are at the point are sustained." R4 *"I think that culture and the history, as well as regulations, collaboration, and legacy systems, are reasons why governments are lagging behind since this is usually burdened for governments to implement their projects in order to be agile (e.g., one reason is funding). One more is that governments often are having technical requirements that businesses do not have for example: having to build websites which are much more accessible to a wider variety of browsers. Governments need to worry about that. I know that in the US, there are certain requirements for governments." R6*	**Lagging:** Funding Organizational / Managerial Factors		
"And I think they have now within public administration – the view is that this is a very strict and conservative commission." R5	**Lagging / Importance:** Leadership		
"In the scoping phase, I think there were at least half a dozen calls and a lot more emails and messages with someone at the partnering institutions who leads the project. That was very helpful. That was not so labor intensive, we just had to get in touch now and then and see how progress is going – With definition the goals and see if the data is available. There was support of DSSG side, and they had a lot of experience what the problems might occur. *In the next phase, there is a lot of technical work that needs to be done to figure out (here we found out who rejects vaccinations.) So, there is a need to generate predictions from the data that is available. So here the DSSG fellows and technical mentors helped a lot. If you summarize DSSG that would be the crucial point – it is the teams of aspiring data science persons. So, there was a lot of support – and it did not just support it was kind of full-time work, and we would have weekly calls to see how it is going. Besides that, there was work on the partnering side from the different stakeholder. It is harder to automatically be ready because you do not know what the partner needs, but there was a willingness to help from DSSG side.* *During the prototyping phase, most of the technical side is on the DSSG side. When it comes to planning and dealing with organizational hurdles, I think there is – ok when we are talking about governmental institutions – most of that will have to remain on the partner side. DSSG can and help with thinking and how to adjust the technical side of those but there will be a lot of things to figure out, and I think the partners need to do it.* *For the implementation phase – I can speak now for my project. The DSSG is very interested, and we were talking about working together on the implementation. The partners have the responsibility to implement it. If they can and want to. Here, the DSSG was very interested in pushing it forward. We are still talking about how else can the DSSG be part of the implementation. It is a different thing to transfer it. Maybe the partner institution does not have the capacities or work with different technologies. So, it was helpful on DSSG side after the project is officially done,*		**Support with:** **1. Scoping: Organizational / Managerial Factors** **2. Prototyping: IT Infrastructure** **Data** **Employees Competencies /** **Organizational / Managerial Factors** **3. Implementation: Organizational / Managerial Factors**	4

there is someone who commits time to walk the partner through this. So that it was is currently happening, and now we are talking about what will happen." R3 *"For the scoping, we require them to admit their application, we have multiple calls, they have to provide sample data, and we try to force them to talk about how they would use the work if it is going to be successful. We try to access their buy-in and their ability. We also try to set expectations for that summer. After scoping and we are starting the prototyping we want side visit at the beginning, and we want one weekly call and typically also answers on data questions and to talk about general challenges with the project. The main responsibility of the implementation lies within the organization itself. For some of the project's fellows have remained involved. (e.g., with NY project). Sometimes DSaPP: takes over and provides full-time persons to work there after the summer. It is the Center for Data Science and Public Policy – that is the center of the University of Chicago who works with the DSSG)" R6*			
"The DSSG gave us access to very skilled people for a reasonable price." R1 *"By the time you managed to do one change, things moved on already – we are always playing catch up. That is again one reason while partnering with an outside organization like the DSSG is an advantage since we are getting access to people with the right skills, are used to have the right tools." R1*	Employees Capabilities Funding Time Management Collaboration / Partnerships Tool Sophistication	**Employees Competencies** **Analytical Tools / IT Infrastructure**	
"Within the first summer, we had a large project which we worked on. When we collaborated with DSSG, that was essentially doubling or triplicate the resources that we had available within the prototyping phase. We handed some stuff to the fellows in Chicago (e.g., Joe Walsh). They helped us tighten up the research design, find resources, datasets, etc. as well as did a lot of the analysis. They were tremendously helpful within this year. From the beginning – scoping, planning, etc. actually executing the work and we brought together a final product after the fellowship even if the results were not what we expected (The publication we jointly produced)." R2	Research Design Organizational Tasks Sources Analytical Skills	**Employees Competencies** **Organizational / Managerial Factors**	
"The data/analytical capability from the project I worked on worked quite good. Most of the capability on the data science part was done by the DSSG team during the prototyping, but there was still technical work to do to make this work of the DSSG available. E.g., anonymizing it before the DSSG can see it (because that is required by law), using SQL, but just from a technical perspective – there are other parts that need to be considered in the prototyping phase." R3	Employees Capabilities Analytical Skills	**2. Prototyping: IT Infrastructure** **Employees Competencies**	
"Further, breaking down barriers would help. Not just the GDPR but also clarify ethical concerns, facilitate an easier automated decision-making process, etc. So, the DSSG helped to demonstrate that the world is not going to end and that algorithms to help out get resources more efficiently. You need to be able to show the policymaker, e.g., that you are able to make a difference and that you can serve the society better with algorithms." R1	Sense of Economic Potential / Awareness Efficiency	**Organizational / Managerial Factors**	
"There is an impact of DSSG to learn, and build-up resources. I think the prototyping phase and the outcomes went quite well (for my project)." R3	**Additional Information**		
"The main idea to start the project was the idea that hopefully there will be awareness raised for this kind of work. Hopefully, people would try data a bit more flexible within their work. In terms of public organizations, we are not flexible. Reconfiguring resources and being dynamic in the changing landscape is difficult, as governmental institutions are, e.g., locked in procurement regulations." R1	**Awareness Flexibility (Peoples Resources and Management)**		**5**
"I think when we were conceptualizing this project, it was not our ambition to automate the decisions. So, in our case, it was more supplemented information helping to assist persons to give good advice for the society. It would certainly enhance their capabilities as a career advisor rather than replace them. That would be our ambition. I would not say that there is currently a change in the managerial factors, but it would hopefully improve the services." R1	**Information Sources Employees Competencies / Skills / Knowledge Services**		
"So, there was a hand-off that was incomplete at that time, but it gives us some ideas of what might be some of the predictive elements that will be in the model" R2	**Analytical Skills**		
"One of the things that helped me. In addition to the work itself, it opened my eyes as far as potentials in terms of approaches to answering questions that were essential to us. I have my own box of tools that I developed over the years from an analytic perspective. But running through a data science project increases awareness about the possibilities in general. Other ways you learn how to approach different things. You also learn to know about machine learning and other tools. For me, it was personal learning. It is highly beneficial for me." R2	**Sense for Economic Potential / Awareness Employees Capabilities**		
"We changed our practices as results with just looking at changes in our client base and then doing some analytics on publicly accessible databases to try and identify where we can go after the high-risk clients with better efficiency – that is an example where we shifted our operations based on analytics to drive processes. *Its an enhancement of the work they are doing its not a substitution of people. We also look at outcomes themselves to get a better outlook about what they are doing in delivering services to maintain high efficiency. So, nurses sometimes need to shift their work." R2*	**Decision Support Efficiency**		

"When implemented, it is much easier for businesses to stretch and align everything between departments and to have data science as a base for all decisions." R3	**Alignment of Resources Alignment of IT/Data + Business**		
"I would rate the ability to learn, reconfigure or build new assets as not really high. I have seen in unrelated projects that there is a willingness to learn from the IT/analysis/tech team of the institution (e.g., with workshops) which is especially important for the implementation phase in data science projects. But it should happen before the implementation with the purpose of implementation. Because you need to have these abilities already to make a decision. I have noticed that there are working and are curious. E.g., I shared some insights about technical tools, and people are excited about it, some of them even used the tools on their own within their free time. There is enthusiasm from the analysis side of people. On the other hand, there are not many opportunities to jump very forward. It is like they are kind of soaked with their work there already. It would be super hard for them to, e.g., attend an intensive course, e.g., one month which would help to drastically advance their capability." R3	**Sense for Economic Potential / Awareness Employees Capabilities**		
"If the goal of the project is to enhance their decision making, the DSSG/data science help how to do it more efficient, while the resources stay the same (not a new intervention per se) but the way they are going to do it after the data science project changes. The role of the leader will probably stay large in other parts of this intervention, but I think if it is the same direction but making it more efficient, it is more realistic that it will happen because of the organizational structure. And if it is discovering some new things that are helping some outcomes the organization cares about, then it is harder. It requires more of a change in the organization. Even if the team leader wants to implement it and change the system, there is someone above that person, and someone above that person, leading to the difficulties of governmental institutions. But it is a positive change since its evidence-based and not based on one specific person's decision. But if the project is something completed different to what they are doing, I think it is hard and a bigger change. Sure, it is a big change, listening to the outcomes of the data and what it is providing." R3	**Dynamic capabilities Leadership Evidenced Based Decision-Making**		
"There was a failure, but I see them as lessons learned and how to go forward to drive ourselves to success. So that in agile, in the case of any failure we try to go to drive in the direction where we can succeed. We do not just stop and give up – we see what we need to bridge this gap in either information, knowledge, technology, human resources, skill sets, policies… there are a lot of things that need to be in place for innovation to go forward a straight path. – and it is never a straight line!" R4	**Learning Persistence Innovation Process**		
"Yes, there is a huge impact while running through the process. I think when you look at the goals, we had set forth leaning about an API and standardizing it, so little work within another health system. Creating real-time interfaces between a public health department and clinics. Bringing in all data that's serrating it into several siloed databases – warehouse we have created for all the data we use for the specific project. In general, we learned a lot of things that help us to go forward. In scoping, sustainability, operationalization, etc." R4	**API Standardizing Processes Sustainability**		
"The timing was right. Now it is less ad hock, it is more internalized and therewith more purposeful. I think we plan for it, allocate, we commit resources. It is a start. The city has two data scientist, and they are not in our department (department of innovation). So, we are reliant on internal partners. We have epidemiologists that are well trained in statistics but not in data science. The workforce needs to be trained in data science methods, and that would be a upskill in the future." R4	**Standardizing Processes Alignment of Resources Partnerships**		
"I hope at least some of them would start thinking about it, about having their own data analyst / chief officers, but perhaps some of the more technical levels, which aids administrative databases, they might be willing to perform some tasks related to data science." R5	**Sense for Economic Potential / Awareness People Resources**		
"So, when we started this, we got the impression that perhaps there will be some data science / AI projects within those organizations, but we were not aware of them, and I think no one was aware of that. But when I started contacting some of these organizations, I realized that there were already some projects being developed within them." R5	**Communication Sense for Economic Potential / Awareness**		
"Yes, there is a positive impact while executing a data science project. Even if you could say this specific project was a failure, the next project could end up having implemented because of the other project. E.g., a project in 2014, there was one project where we did not really have the data that we needed to do that project correctly, and they went back, collected data and did a project in 2015, and it was a success even in the implementation phase. (…) Even during the scoping phase, we are saying the partners, e.g., 'you are not ready. You need to do x, y, and z before you can do a data science project. And they did x, y, and z and we partner with them later." R6	**Sustainability Sources**		
"Yes, I totally agree with you. There is a positive impact on the organization while executing a data science project. That is something that I can say now that, that this initiatives show is that it opens a lot of minds within public administration to the importance of data science and prototyping. (…) So, I think this definitively has an impact on even the most conservative organizations within the public area (sensitive	**Sense for Economic Potential / Awareness**		

to realizing data, etc.), they are now more open to this because we have started this initiative." R5			
"Database administrators could be helpful to implement projects for people who can learn under staff. Even if some partners have good people, so many don't even know what we are doing. And the partners who participated twice within the fellowship showed the ability to learn (e.g., how to build a model) as they were able to do it on their own from there on." R6 *"Two is to set up some kind of field trial with that way you have evidence. That should be an ongoing process and it probably not discussed enough. Even if it is a field trial, people look at it as a one-off, but things change over time (computers might get better, or worse compared to people) so at least some of the selection and model evaluation should take place randomly so we can see what the effect is." R6*	**Sustainability**		
"I think it would be difficult, to view that some organizational structures would change, because of these projects. At least at this point. For one, it is difficult to get the right skills in public administration for maintaining these projects, I think they are becoming more aware – perhaps, one or two would be willing to hire some data scientists/chief data officer, they would perhaps improve a little bit their data structures and would improve the access mechanisms, but right now – in my opinion – they will be no major changes within the organizational structure because of this. I think one or two would create a small innovation hub/center/ team – but even then, I would think it would be very difficult because there are a lot of constraints on hiring people regarding financial resources, etc." R5	**Sense for Economic Potential / Awareness** **People Resources** **Accessibility / Sharing**		
"I think that the outcomes can give the workforce different roles, which could be an advantage, as data is the best choice for decision-making. However, a human component is especially important for creativity." R6 *"organizational change is not necessary, but it provides toolsets that allow the programs, supervisor or manager for top data-driven decision abilities - so, it is complementary." R4*	**Peoples Resources**		
"But regardless of the outcome, if a project failed or succeeded within one step, there are positive externalities. Part of it is even being aware of what can be done. E.g., people get excited about that they can build a model that is updated quite often, to connect it to the dashboard, for example. Being aware that it can be done. It is helpful because it makes the team more ambitious at a more tangible level. In my case, it was inaccessible silo data as a result of the project which were brought together to a central source, and now, they can work with it. And maybe for the future, they will have an easier time to integrate the data that they need." R3	**Sense for Economic Potential / Awareness** **Accessibility / Sharing** **Integration**		
"Ideally, a project is successful when it is implemented, but here it is about reaching the next phase; plus, several different things could mean success: the help for the social good, inspiration of people for the future, training of the partner and fellows, etc. Not all projects need to possess all the given factors in order to be successful in some kind" R6 all	**Success of a project**	**Additional Information**	
"When nothing is implemented, or no new projects got inspired, partner did not earn additional resources" R6	**Failure of a project**	**Additional Information**	

R1: Jamie Stainer, Principal of the Youth Training Statistic and Research Branch, Department for the Economy of the UK
R2: Bill Thorland, Director of Evaluation and Research, Nurse-Family Partnership
R3: Tin Oreskovic, Partner for Croatian Institute of Public Health
R4: Raed Mansour, Director, Chicago Department of Public Health
R5: Rui Lourenco, Technical Advisor, Republic of Portugal
R6: Joseph Walsh, Data Scientist, Center for Data Science & Public Policy & DSSG

E. Network of Resources and Capabilities

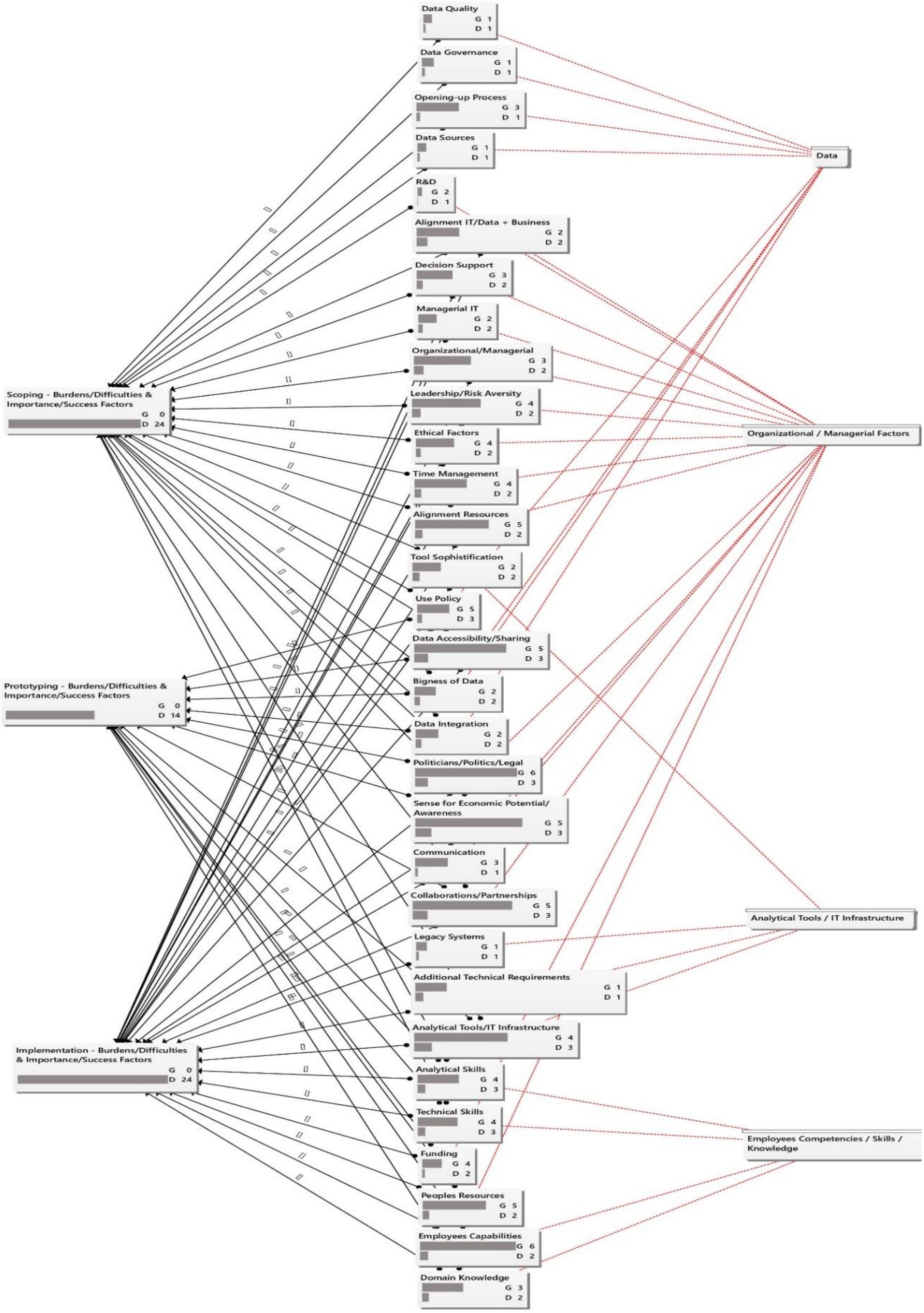

Figure 7: Importance, Relation & Frequency of Resources and Capabilities: The figure reveals all substantial resources and capabilities for data science projects in governmental institutions reported by the interviewees. All four categories are related to all three phases, while, e.g., especially 'data' related factors are critical within the scoping phase and 'employees' competencies/skills/knowledge' indicate a high influence of success for the implementation phase of a data science project.